Network Structure Identification using Corrupt Data-Streams

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Venkat Ram Subramanian

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Advisor: Andrew Lamperski, Co-Advisor: Murti Salapaka

August, 2021

© Venkat Ram Subramanian 2021
ALL RIGHTS RESERVED

Acknowledgements

This milestone was made possible by a network of people and circumstances.

I will always be grateful for the summer of 2015 when I started working on my first research problem with Prof. Andrew Lamperski. The joy of problem formulation and solving coupled with Andy's zeal for research ignited a spark within me which led to my decision to pursue a doctoral degree.

I am highly indebted to my Ph.D. advisors, Prof. Andy Lamperski and Prof. Murti Salapaka, for their insight, knowledge, encouragement, and guidance that were crucial to my growth and development throughout graduate school. They have been instrumental in shaping my capability to conduct scientific research. Under their mentorship, I have developed considerable patience, ability to think critically and creatively, and the confidence to approach and solve problems. I hope to carry these values through the rest of my career.

I sincerely thank Dr. Deepjyoti Dekha, Scientist at Los Alamos National Laboratory, and Dr. Saurav Talukdar, Engineer at Google, for their time and efforts in shaping the exact network structure learning component of this dissertation. Their knowledge, experience, and perspectives led to several invigorating discussions which taught me so much.

For all the fun brainstorming sessions, valuable feedbacks, and most importantly for being the best peers I could ask for on this journey, I would like to acknowledge all my lab mates – Dr. Jianjun Yuan, Dr. Bolei Di, Tyler Lekang, Steve Thomas, Harish Doddi, Dr. Sourav Patel, Rachit Shrivastava, Sivaraman Rajaganapthy, Mishfad Veedu, Dr. Govind Saraswat and Vivek Khatana.

Through good days and bad my constants have been Dr. Abhinav V. Sambasivan and Rohit Sridhar. I thank them for being terrific friends, roommates, and providing

limitless fun and humor. I would also like to thank Ashwin, Ramesh, Dinesh, Ramya, Hariharasudhan, Karthik, Subhash, and Deepak for making my life outside Ph.D. full of happiness. Graduate school and life in Minneapolis will be an unforgettable experience because of you all.

I also express my gratitude to my cousin Meena, and her parents (Lalitha and Easwaran), for encouraging and backing my decision to pursue a doctoral degree. I would like to thank my parents-in-law, and sister-in-law (Mathangi) for cheering me on.

I am forever indebted to my parents, Lakshmi Subramanian and Subramanian Raman for their faith in me, unyielding support, and blessings. Their happiness and pride at seeing me become the first doctorate in the family is overwhelming. I am also grateful to my grandmother Savithri Venkatesh, my sister Sudha, my brother-in-law Venkat Giridhar for their relentless encouragement and care.

Last but not the least, I thank my lovely wife Dr. Vaishnavi Soundararajan, who has been tremendously supportive and caring. Her positive attitude and animated personality are inspiring and have always lifted my spirits whenever I felt low. I am grateful for her companionship, for being in my life, and for making me a better person every day (also for editing this acknowledgement!).

Dedication

Sri Rama Jayam

To

My parents Lakshmi Subramanian and Subramanian Raman

My grandparents Smt Savithri Venkatesh, Shri N.V. Venkatesan, Shri K.S. Raman, and Smt S. Lakshmi

Bhagwan

Abstract

Many complex systems lend themselves to effective modeling described by a network of dynamically interacting agents. Such modeling is prevalent in many application domains that include climate science, neuroscience, internet-of-things, power grids, and econometrics. The evolution of these systems is governed by the interdependencies and interactions between the agents that can contain feedback loops. Identification of the presence or absence of influence pathways among the agents is of primary importance that enables subsequent analytics in networked systems such as identifying central agents and clusters, devising control strategies in distributed systems, and resource allocation. In most application domains, the nature of the relationships and interdependencies cannot be easily modeled using first principles. Furthermore, in many such systems, it is not possible to deliberately affect the system, and thus passive or noninvasive methods are required. The existing methods of network identification do not account for the common ways through which data gets corrupted. In real-world systems, sensor readings can be inaccurate, clocks can get out of sync, and messages can get lost in transmission over a wireless network. The focus of this research is to incorporate realistic modeling assumptions on data streams and characterize the effects of data corruption on network identification using passive means.

We show that identifying the structure of networked systems using corrupt measurements results in the inference of spurious links. The effects of data corruption on network reconstruction are characterized with provable guarantees on the quality of construction with respect to the generative models considered. A wide range of generative models that underlie the data streams are considered that include static interactions (Markov random fields), linear time-invariant dynamical systems, and nonlinear dynamical models. We examine both causal and non-causal inference methods. In both cases, we provide an exact characterization of the location of spurious links. Our results show that the spurious links are localized to the neighborhood of the corrupted node. All our solution methodologies utilize only the time-series observations without any knowledge of the system parameters.

Our precise characterization of the erroneous links is further exploited when the network has special structural properties. There are several physical systems, especially flow-driven systems like power grids, heat transfer networks, and fluid flow networks, where every dynamic coupling between the agents/nodes is bi-directional. In such systems, identifying unidirectional links in reconstruction lead to the conclusion that such links arise from data corruption. We utilize our precise characterization of spurious links to detect and localize all corrupt nodes in the network. It is imperative that learning the exact network representation of such systems without spurious links is needed for performing accurate state estimation, control, and optimization. To this, we developed methods to remove all spurious links and identify the exact structure of bi-directed networks despite of data corruption.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Why Networks?

Models of complex systems as networks of interacting systems are ubiquitous across many domains such as climate science [1–3], epidemiology [4, 5], neuroscience [6–8], metabolic pathways [9–11], quantitative finance [12–14], the internet-of-things (IOT) [15–17] and video streaming [18]. In many of these systems such as brain, communication network [19] and power grid [20, 21] there is a clear physical network of interacting agents. Other applications such as social networks [22] and quantitative finance are examples of networks of nonphysical systems. In all the above domain examples, identification of influence pathways in networked systems is a necessary first step that enables subsequent analytics such as: identifying important agents and clusters [23,24], devising control strategies in distributed systems to steer the system toward the desired behavior [25–28] and resource allocation [29, 30]. The need for frameworks for analyzing and designing networks of dynamical systems has received renewed emphasis fueled by new technologies and paradigms. For example, in the internet-of-things (IOT) (see [31, 32]), data collected from ubiquitously sensorized devices and/or from many sensors of a single large system (such as an aircraft) are processed to glean important insights and devise control strategies. Here, the data streams being sensed can be dynamically related, where the interdependency can be caused by the interaction physics of system components.

1.2 Challenges in Network Identification

Approaches for network structure identification may employ either *active* [33] methods, or *passive* [34] methods. The active method of network identification involves removal of nodes in the network, and/or external excitation/injection of signals to evaluate the effects in the network [35]. However, in scenarios such as the power grid, climate science, and financial markets it is impractical, impossible, or impermissible to externally influence the system. Here, network structure identification must be achieved via non-invasive or *passive* means. The passive identification of a network of dynamically related agents is becoming more viable with sensors and measurements becoming inexpensive coupled with the ease and capability of communicating and processing information.

In an effective approach, the interconnectivity structure of a complex networked system is modeled as a graph where nodes represent individual entities and links represent a notion of coupling or dependence among them. Initial efforts in graphical model identification are restricted to modeling the interactions between the time-series data as static relations [36, 37]. Limiting to static-relations are significantly restrictive in networks of dynamical systems where at any given time, the present value of an agent may depend on the past values and the present value of other agents (also including itself) and often feedback loops of dynamical interactions between the agents are prevalent. This poses further challenge in establishing guarantees for consistency of the inferred network structure.

In most applications, a significant challenge is that the measurements of data streams are plagued by uncertainties [38,39]. Interconnected networked systems are not immune to asynchronous sensor clocks [40–46]. Other prevalent forms of measurement uncertainties include inaccurate sensor readings caused by measurement noise [47–52], and messages getting lost due to packet drops [53–57]. Additionally, such measurement channel uncertainties can be temporally correlated. It is also possible that such forms of data imperfection can be introduced deliberately; as malicious attacks [58–62]. Here, distinguishing corrupted data from good data is challenging and often intractable. It is conceivable that errors in the inferred network structure plague implementation strategies for estimation, control and optimization in networked systems. Thus, for any framework for the identification of influence pathways it is imperative that the effects

of uncertainty in the dynamically related data streams being processed be explicitly quantified and understood.

1.3 Thesis Contribution

Network identification for dynamical systems is extensively studied. See [35, 63–82, 82–93]. Sparsity constrained methods are studied in [63–69], where the sparsity of the parameter estimates are enforced to reflect the absence of the corresponding edges in the network. However, this requires the selection of regularization parameter and the consistency of the structure estimates depends on the chosen parameter. Furthermore, reconstruction guarantees using uncertain measurements are not addressed. Active methods of network identification are studied in [35, 70–72]. Here too, the data measurements are assumed to be perfect without any distortions. Passive methods of network reconstruction are extensively studied [73–89, 126]. However, measurement data uncertainies and their effects on the quality of reconstructed network have not been addressed. Another theme of research address the question of identifiability [82, 90–93], under what conditions is the network reconstructed from the measured data unique? Here too, the data-streams are assumed to be perfectly measured without uncertainties and disturbances.

In summary, though network identification is extensively studied, and despite its significance, there is a gap in addressing the effects of measurement uncertainties on network identification. This dissertation tackles the problem of identifying the interconnection structure in networks of dynamical systems with corrupted data streams using passive methods. We show that identifying the structure of networked systems using corrupt measurements results in the inference of spurious links. We provide a rigorous characterization on the location of spurious links. The effects of data corruption on network reconstruction are characterized with provable guarantees on the quality of construction with respect to the generative models considered. A wide range of generative models that underlie the data streams are considered that include static interactions, linear time-invariant (LTI) dynamical systems, and nonlinear dynamical models. Rigorous data corruption models are developed that cover a broad class of practically relevant measurement signal disturbances including nonlinear dynamics and

temporal correlations. We do not use any information about the system parameters or inputs, and utilize only the time-series measurements. The specific contributions of the dissertation are summarized as follows:

- Considering network reconstruction for linear time-invariant systems admitting non-causal dynamics, we show that the spurious links are characterized by the appearance of cliques that are localized to the neighborhood of the corrupted node. We provide a precise characterization of the location and pattern of spurious links arising due to multiple corrupt nodes. Our results on uncertainty propagation during network reconstruction can be applied to guide strategic placement of high-fidelity sensors and network synthesis that are resilient to and minimize the effects of data corruption. This is significant during planning stage and deployment of monitoring devices in applications such as power distribution [94], water distribution [95], and structural health monitoring [96, 97], where network topology is first identified using sensor measurements for detecting faults and outages.

- Next, we focus on networks governed by static interactions and show that spurious links arising from data corruption is a more general phenomenon. Here, the interactions though static, admit nonlinear dependencies. The graphical models describing such networks are called *Markov random fields* (MRF). MRF framework is prevalent across several application domains that include gene regulatory networks [10,138], social networks [139], statistical physics [140] image processing [98, 141, 142], robotics [99, 143], finance [144], and climate science [3, 145, 146]. Structure inference of MRF is extensively studied [36, 37, 101, 102, 147–157, 157–173]. However, state-of-the-art methods do do not adequately address how corrupted observations affect structure inference of MRF (see Section 3.1 for a detailed overview on related work). We establish the extent of uncertainty propagation of spurious links in inference of MRF without requiring any structural assumptions on the underlying graph nor any prior knowledge on the system nor the noise parameters.

- We provide necessary and sufficient conditions that delineate the effects of data corruption on the directed network structure inference for nonlinear dynamical systems. We present a tight characterization for the spurious links that arise due to

data corruption by determining their location and orientation. The knowledge of the directionality of spurious effects can lead to a better assessment of the quality of reconstruction of mutual influences, and can aid elimination of spurious links using complimentary methods. Furthermore, the knowledge of causal influence structure is required a priori to perform local module system identification in networked systems [103–114], where a local part of the network is of interest and needs to be estimated. Thus, our results serve as a necessary first step in understanding what part of network reconstruction can be trusted to facilitate accurate system identification.

- We show that our precise characterization of the spurious links inferred can be further exploited when the network has special structural properties. There are several physical systems, especially flow-driven systems like power grid [115–117], fluid flow [118–121] and heat transfer networks [122–125], where every dynamic coupling between the agents/nodes is bi-directional. In such systems, identifying unidirectional links in reconstruction lead to the conclusion that such links arise from data corruption. Using our precise characterization of spurious links, we develop a procedure to detect the location of all corrupt nodes in bi-directed networks.

- Next, we consider bi-directed networks of linear time-invariant dynamical systems [107, 113, 126–128]. We develop an algorithm to remove all spurious links and determine the exact network structure of bi-directional linear time-invariant systems despite of data corruption. This is significant for adopting the *errors-in-variables* method [104, 129–132] to perform accurate system identification in networked systems in presence of measurement noise. Errors-in-variables framework requires the knowledge of exact network topology prior to estimation of transfer functions. Thus, our results serve as a necessary first step to determine the exact network topology despite of data corruptions and complimentary methods using errors-in-variables approach can then be applied to perform system identification.

1.4 Thesis Outline

- Chapter 2 presents a wide range of perturbation models for signal disturbances and characterizes the effects of data corruption on network reconstruction for linear time-invariant systems.
- In Chapter 3, we present the results for inferring spurious links in networks admitting static and nonlinear dependencies, that is, the time series are modeled as random variables.
- Chapter 4 considers networks of nonlinear and causal dynamical interactions. For a more general class of perturbation models, an exact characterization on the location and orientation of spurious links in the inferred graph is provided.
- The conditional directed information estimator for directed network inference is developed in Chapter 5. The consistency of the estimator is also discussed here.
- Corruption detection algorithm in networks of bidirectionally coupled nonlinear dynamic influences is developed in Chapter 6.
- Chapter 7 presents the algorithm to determine the exact network representation of bi-directional linear time-invariant systems using corrupted data streams.

Chapter 2

Network Structure Identification in LTI Systems using Corrupt Data Streams

2.1 Introduction

As mentioned before, the interconnectivity structure of a complex networked system is effectively modeled as a graph where nodes represent individual entities and links represent a notion of coupling or dependence among them. In this chapter, we focus on networks where every dependence between the agents is governed by linear time-invariant dynamics. Network identification for linear systems has been extensively studied. Below, we will present an overview of several research themes in linear system network identification. However, the majority of works assume that the measurements are perfect.

2.1.1 Related Work

Identifiability conditions for determining the transfer functions are provided in [82]. It is shown that a network is identifiable if every node signal is excited by either an external input or a noise signal that is uncorrelated with the input/noise signals on the other nodes. The effects of data corruption are not studied in this work.

For partially observed states, authors in [103] provide necessary and sufficient conditions for *generic* identifiability of all or a subset of the transfer functions in the network. Similarly, the notion of *global* identifiability has been studied in [105]. However, in both the articles, the topology of the network is assumed to be known a priori. Moreover, data measurements are assumed to be perfect.

The problem of learning polytree structures has been studied in [80] and [85]. The authors provide guarantees of a consistent reconstruction. However, the class of network structures was restricted to trees and the data measurements are assumed to be ideal. In this work, we make no such assumptions on network structures and we study the problem when time-series data measurements are imperfect.

Authors in [74] leveraged multivariate Wiener filters to reconstruct the undirected topology of the generative network model. With assumptions of perfect measurements, and linear time invariant interactions, it is established that the multivariate Wiener filter can recover the *moral graph*. In other words, for each node, its parents, co-parents and children are detected. Here, even without measurement and communication uncertainties, spurious influences are detected; however, they remain local and within one hop of a true interaction. It is imminent and desirable to establish conditions under which the combination of spurious links from the method of [74] and the effects of data corruption do not compound so that negative effects remain localized.

The authors in [133], [134] use dynamical structure functions (DSF) for network reconstruction [135] and consider measurement noise in the network dynamics. The proposed method first finds optimal DSF for all possible network interconnection structures and then adopt a model selection procedure to determine the best estimate. The authors concluded that the presence of noise and non-linearities can even lead to spuriously inferring fully connected network structures. Also, the authors concluded that the performance of their algorithms degrades as noise and network size increase. However, a precise characterization of such spurious inferences in structure was not provided.

[76] and [136] identify causal dependencies in network of LTI systems driven by unknown intrinsic noise inputs. Assuming that the network structure is known, *errors-in-variables* framework for system identification with additive sensor noise is studied in [129], [104]. However, in this work we do not assume that the interaction structure is known.

2.1.2 Our Contributions

In this thesis, our problem of interest is to determine the Boolean structure of a network, that is, identification of presence or absence of links, using passive means from corrupt data-streams, and characterize the spurious links that can appear due to data-corruption.

In order to rigorously model data corruption, we present a general class of signal disturbance models based on randomized state-space systems. This class of disturbances subsumes many uncertainties that are prevalent in applications. We provide a detailed description on how the corruption model affects the second order statistics of the data-streams.

We present the results for inferring the network topology for LTI systems from corrupt data-streams. Specifically, we identify a set of edges in the network in which spurious links could potentially appear. The results can be utilized to understand what part of the reconstruction can be trusted and to allocate sensor resources in order to minimize the effects of data corruption. This accentuates the pertinence of the methods for wider class of networks that has feedback loops and allows instantaneous and/or non-causal dynamic relations.

We start by reviewing earlier work on LTI network identification using power spectra in Section 2.2. In Section 2.3, we describe our data corruption models. In Section 2.4, we characterize the spurious links due to data corruption for LTI systems. Simulation results are provided in Section 2.5.

Notation

Y denotes a vector with y_i being i^{th} element of Y.
$z_i[\cdot]$ denotes a sequence and $z_{i,t}$ denotes $z_i[t]$.
$\|\cdot\|$ denotes standard Eucledian norm for vectors.
P_X represents the probability density function of a random variable X.
$X \perp\!\!\!\perp Y$ denotes that the random variables X and Y are independent.
$i \to j$ indicates an arc or edge from node i to node j in a directed graph.
$i - j$ denotes an undirected edge between nodes i, j in an undirected graph.

If $M(z)$ is a transfer function matrix, then $M(z)^* = M(z^{-1})^T$ is the conjugate transpose.
$\mathbb{E}[\cdot]$ denotes expectation operator.
$R_{XY}(k) := \mathbb{E}[X[n+k]Y[n]]$ is the cross-correlation function of jointly wide-sense stationary(WSS) processes X and Y. If $Y = X$ then $R_{XX}(k)$ is called the auto-correlation.
$\Phi_{XY}(z) := \mathcal{Z}(R_{XY}(k))$ represents the cross-power spectral density while $\Phi_{XX}(z) := \mathcal{Z}(R_{XX}(k))$ denotes the power spectral density(PSD) where $\mathcal{Z}(\cdot)$ is the Z-transform operator.
b_i represents the i^{th} element of the canonical basis of $\mathbb{R}^n$.

2.2 Background on LTI Network Identification

This section reviews earlier results on network identification from ideal data streams. See [74]. Required graph theoretic notions are described in Subsection 2.2.1. The formal model of networked LTI systems is presented in Subsection 2.2.2. Then, a result on network identification via power spectra is given in Subsection 2.2.3. In later sections, we will analyze these results in the case that data has been corrupted.

2.2.1 Graph Theoretic Preliminaries

We will review some terminology from graph theory needed to describe the background results on LTI identification. For reference, see [36].

Definition 1 (Directed and Undirected Graphs)**.** *A* directed graph *G is a pair (V, A) where V is a set of vertices or nodes and A is a set of edges given by ordered pairs (i, j) where $i, j \in V$. If $(i, j) \in A$, then we say that there is an edge from i to j. (V, A) forms an* undirected graph *if V is a set of nodes or vertices and A is a set of the un-ordered pairs $\{i, j\}$.*

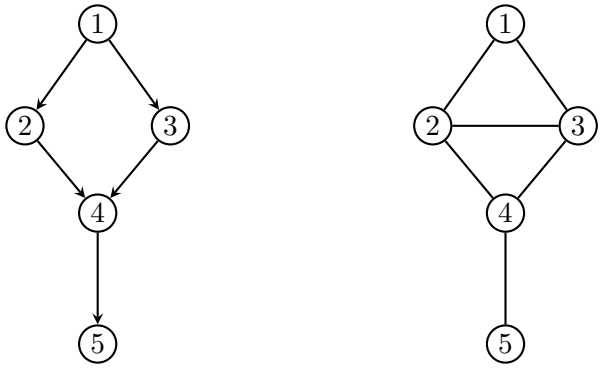
We also denote an undirected edge as $i - j$.

Definition 2 (Children and Parents)**.** *Given a directed graph $G = (V, A)$ and a node $j \in V$, the children of j are defined as $\mathcal{C}(j) := \{i | j \to i \in A\}$ and the parents of j as $\mathcal{P}(j) := \{i | i \to j \in A\}$.*

Definition 3 (Kins)**.** *Given a directed graph $G = (V, A)$ and a node $j \in V$, kins of j are defined as $\mathcal{K}_j := \{i | i \neq j \text{ and } i \in \mathcal{C}(j) \cup \mathcal{P}(j) \cup \mathcal{P}(\mathcal{C}(j))\}$. Kins are formed by parents, children and spouses. A spouse of a node is another node where both nodes have at-least one common child.*

Definition 4 (Moral/Kin Graph)**.** *Given a directed graph $G = (V, A)$, its moral-graph is the undirected graph $G^M = (V, A^M)$ where $A^M := \{\{i, j\} | j \in V, i \in \mathcal{K}_j\}$.*

Fig. 2.1 provides an example of a directed graph and its moral graph.



(a) Directed Graph (b) Kin Graph

Figure 2.1: 2.1(a) Directed Graph and 2.1(b) its moral/kin Graph.

2.2.2 Dynamic Influence Model for LTI systems

Here the *generative model* that is assumed to generate the measured data is described. Consider N agents that interact over a network. For each agent i, we associate an observable discrete time sequence $y_i[\cdot]$ and a hidden noise sequence $e_i[\cdot]$. The process $e_i[\cdot]$ is considered innate to agent i and thus e_i is independent of e_j if $i \neq j$. We assume e_i and y_i to be jointly wide-sense stationary stochastic processes. In particular, we assume they are bounded in a mean-square sense: $\mathbb{E}[\| y_i[t] \|^2] < \infty$ and $\mathbb{E}[\| e_i[t] \|^2] < \infty$.

Let Y denote the set of all random process $\{y_1, \ldots, y_N\}$ with a parent set $\mathcal{P}'(i)$ defined for $i = 1, \ldots, N$. The parent set $\mathcal{P}'(i)$ associated with agent i does not include i. The process y_i depends dynamically on the processes of its parents, y_j with $j \in \mathcal{P}'(i)$ through an LTI filter whose impulse response is given by $\mathcal{G}_{ij}$. Therefore, dynamics of

node i takes the form:

$$y_i[t] = \sum_{j \in \mathcal{P}'(i)}^{N} (\mathcal{G}_{ij} * y_j)[t] + e_i[t] \quad \text{for } i = 1, \ldots, N. \tag{2.1}$$

where $*$ denotes convolution operation. Performing a Z-transform on both sides gives

$$y_i(z) = \sum_{j \in \mathcal{P}'(i)}^{N} \mathcal{G}_{ij}(z) y_j(z) + e_i(z) \quad \text{for } i = 1, \ldots, N. \tag{2.2}$$

For compact notation, we will often drop the z arguments. Let $y = (y_1, y_2, \ldots, y_N)^T$ and $e = (e_1, e_2, \ldots, e_N)^T$. Then (2.2) is equivalent to

$$y = \mathcal{G}(z) y + e. \tag{2.3}$$

The diagonal entries $\mathcal{G}_{ii}(z)$ are considered to be zero. We refer to (2.3) as the Dynamic Influence Model (DIM). Here, $\mathcal{G}$ is termed as the DIM *generative connectivity* matrix. The DIM will be denoted by $(\mathcal{G}, e)$.

Remark 1. *The process noise in* (2.1) *can be correlated across time. In that case, e_i is assumed to be represented as the convolution of white noise with a stable LTI filter.*

Remark 2. *The diagonal entries, $\mathcal{G}_{ii}(z)$ are considered to be zero only for simplification purposes to remove self-dependence in the dynamics. As will be seen later in sub-section 2.2.3, this enables us to consider Wiener filter projection of signal y_i on all signals except y_i. Moreover, we can model the self-dependence and include it in the DIM through the process noise sequence by convolving a zero mean white noise with $\mathcal{G}_{ii}(z)$.*

We illustrate the notation by an example. Consider a network of five agents whose node dynamics are given by,

$$\begin{aligned} y_1 &= e_1 \\ y_2 &= \mathcal{G}_{21}(z) y_1 + e_2 \\ y_3 &= \mathcal{G}_{31}(z) y_1 + e_3 \\ y_4 &= \mathcal{G}_{42}(z) y_2 + \mathcal{G}_{43}(z) y_3 + e_4 \\ y_5 &= \mathcal{G}_{54}(z) y_4 + e_5 \end{aligned} \tag{2.4}$$

$$\text{with } \mathcal{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \mathcal{G}_{21} & 0 & 0 & 0 & 0 \\ \mathcal{G}_{31} & 0 & 0 & 0 & 0 \\ 0 & \mathcal{G}_{42} & \mathcal{G}_{43} & 0 & 0 \\ 0 & 0 & 0 & \mathcal{G}_{54} & 0 \end{bmatrix}.$$

Definition 5 (Generative Graph)**.** *The structural description of* (2.3) *induces a generative graph* $G = (V, A)$ *formed by identifying each vertex* v_i *in* V *with random process* y_i *and the set of directed links,* A*, obtained by introducing a directed link from every element in the parent set* $\mathcal{P}'(i)$ *of agent* i *to* i*.*

Note that we do not show $i \to i$ in the generative graph and neither do we show the processes e_i. The generative graph associated with the examples described in (2.4) is given by Fig. 2.1 (a).

2.2.3 Identification from Ideal Measurements

The following results are obtained from [74] where the authors have leveraged Wiener filters for determining generative graphs of a DIM.

Theorem 1. *Consider a DIM* $(\mathcal{G}, e)$ *consisting of* N *nodes with generative graph* G*. Let the output of the DIM be given by* $y = (y_1, \ldots, y_N)^T$*. Suppose that* S_j *is the span of all random variables* $y_k[t],\ t = \ldots -2, -1, 0, 1, 2 \ldots$ *excluding* y_j*. Define the estimate* $\hat{y}_j$ *of the time-series* y_j *via the optimization problem of*

$$\min_{\hat{y}_j \in S_j} \mathbb{E}\left[(y_j - \hat{y}_j)^T (y_j - \hat{y}_j)\right].$$

Then a unique optimal solution to the above exists and is given by

$$\hat{y}_j = \sum_{i \neq j} \mathbf{W}_{ji}(z) y_i \tag{2.5}$$

where $\mathbf{W}_{ji}(z) \neq 0$ *implies* $y_i \in \mathcal{K}_{y_j}$ *(equivalently* $y_j \in \mathcal{K}_{y_i}$*); that is* i *is a kin of* j*.*

The solution in (2.5) is the Wiener Filter solution which is given by $\Phi_{y_j y_{\bar{j}}} \Phi^{-1}_{y_{\bar{j}} y_{\bar{j}}}$ where $y_{\bar{j}}$ denotes the vector of all processes excluding y_j and Φ denotes the power spectral density. Thus, Theorem 1 implies that we can reconstruct the moral graph of a DIM by

analyzing the joint power spectral density of the measurements. The following corollary gives a useful characterization of the inferred kin relationships in terms of the sparsity pattern of Φ_{yy}^{-1}.

Corollary 1. *Under the assumptions of Theorem 1, let Φ_{yy} be the power spectral density matrix of the vector process y. Then the (j,i) entry of Φ_{yy}^{-1} is non zero implies that i is a kin of j.*

Remark 3. *$\Phi_{yy}^{-1}(i,j)$ is described by (i,j) entry of $(I-\mathcal{G}(z))^*\Phi_e^{-1}(I-\mathcal{G}(z))$. Specifically, $\Phi_{yy}^{-1}(i,j) = -\mathcal{G}_{ij}\phi_{e_i}^{-1} - \mathcal{G}_{ji}^*\phi_{e_j}^{-1} + \sum_k \mathcal{G}_{ki}^*\mathcal{G}_{kj}\phi_{e_k}^{-1}$ where $k \in \mathcal{C}(i) \cap \mathcal{C}(j)$. For i and j being kins but $\Phi_{yy}^{-1}(i,j)$ to be zero, the transfer functions in $\mathcal{G}$ must be belong to a set of measure zero on space of system parameters. For example, system dynamics with transfer functions being zero or a static system with all noise sequences being identical. Therefore, except for these restrictive cases, the results in Theorem 1 and Corollary 1 are both necessary and sufficient. See [74] for more details.*

2.3 Uncertainty Description

Subsection 2.2.3 describes a methodology from [74] for guaranteed kinship reconstruction based on Wiener filtering. However, the results assume that the signals, y_i, are measured perfectly. This chapter aims to explain what would happen if we attempted to apply the reconstruction method to data that has been corrupted. We will see that extra links appear in the reconstruction, and characterize the pattern of spurious links.

Subsection 2.3.1 presents the general class of data corruption models studied for LTI systems. The modeling framework is sufficiently general to capture a variety of practically relevant perturbations, such as delays and packet loss. However, we will see that all of the corruption models have similar effects on the observed power spectra. Specific examples of perturbation models are described in Subsection 2.3.2.

2.3.1 Random State Space Models

This subsection presents the general class of perturbation models. Consider i^{th} node in a network and let it's associated unperturbed time-series be y_i. The corrupt data-stream u_i associated with i is considered to follow the stochastic linear system described below:

$$x_i[t+1] = A_i[t]x_i[t] + B_i[t]y_i[t] + w_i[t] \tag{2.6a}$$

$$u_i[t] = C_i[t]x_i[t] + D_i[t]y_i[t] + v_i[t], \tag{2.6b}$$

where x_i denotes hidden states in the stochastic linear system that describes the corruption. Here, the matrices, $M_i[t] = \begin{bmatrix} A_i[t] & B_i[t] \\ C_i[t] & D_i[t] \end{bmatrix}$ are independent, identically distributed (IID) and independent of $y_i[t]$. The terms $w_i[t]$ and $v_i[t]$ are zero-mean IID noise terms which are independent of $M_i[\cdot]$ and $y_i[\cdot]$ and have covariance:

$$\mathbb{E}\left[\begin{bmatrix} w_i[t] \\ v_i[t] \end{bmatrix}\begin{bmatrix} w_i[t] \\ v_i[t] \end{bmatrix}^\top\right] = \begin{bmatrix} W & S \\ S^\top & V \end{bmatrix}. \tag{2.7}$$

For distinct perturbed nodes, $i \neq j$, we assume that $M_i[]$, $w_i[]$, and $v_i[]$ are independent of $M_j[]$, $w_j[]$, and $v_j[]$.

Denote the means of the state space matrices by $\bar{A}_i = \mathbb{E}[A_i[t]]$, $\bar{B}_i = \mathbb{E}[B_i[t]]$, $\bar{C}_i = \mathbb{E}[C_i[t]]$, and $\bar{D}_i = \mathbb{E}[D_i[t]]$.

Let h_i be the impulse response of the system defined by $\bar{A}_i, \bar{B}_i, \bar{C}_i, \bar{D}_i$:

$$h_i(k) = \left[\begin{array}{c|c} \bar{A}_i & \bar{B}_i \\ \hline \bar{C}_i & \bar{D}_i \end{array}\right](k) \tag{2.8}$$

Note that $\bar{u}_i[t] = \mathbb{E}[u_i[t]|y_i] = (h_i \star y_i)[t]$.

Theorem 2. *Assume that $M_i[t]$ has bounded second moments and for all positive definite matrices Q, the following generalized Lyapunov equation has a unique positive definite solution, P:*

$$P = \mathbb{E}[A_i[t]PA_i[t]^\top] + Q. \tag{2.9}$$

Define $\Delta u_i[t] := u_i[t] - \bar{u}_i[t]$. Then, the signals u_i will be wide sense-stationary with cross-spectra and power spectra of the form:

$$\Phi_{u_iu_i}(z) = H_i(z)\Phi_{y_iy_i}(z)H_i(z^{-1}) + \theta_i(z) \tag{2.10a}$$

$$\Phi_{u_iy_i}(z) = H_i(z)\Phi_{y_iy_i}(z) \tag{2.10b}$$

where, $H_i(z) = \mathcal{Z}(h_i)$ and $\theta_i(z) = \mathcal{Z}\left(R_{\Delta u_i \Delta u_i}[k]\right)$.

The proof is given in Appendix A.1.

2.3.2 Data Corruption Examples

We will highlight a few corruptions that are practically relevant to exemplify the above model description. More complex perturbations can be obtained by composing these models.

Random Delays

Randomized delays can be modeled by

$$u_i[t] = y_i[t - d[t]] \tag{2.11}$$

where $d[t]$ is a random variable. For example, if $d[t] \in \{1, 2, 3\}$, then randomized delay model can be represented in state-space form with no additive noise terms and state space matrices given by:

$$\left[\begin{array}{c|c} A_i[t] & B_i[t] \\ \hline C_i[t] & D_i[t] \end{array}\right] = \left[\begin{array}{c|c} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} & \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ \hline b_{d[t]}^\top & 0 \end{array}\right],$$

where b_1, b_2, and b_3 are the standard basis vectors of $\mathbb{R}^3$.

Say that $d[t] = j$ with probability p_j, for $j = 1, 2, 3$. Then

$$H_i(z) = p_1 z^{-1} + p_2 z^{-2} + p_3 z^{-3}. \tag{2.12}$$

Let $p = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}$. The formal description to compute the expression for $\theta_i(z)$ is discussed in Lemma 7 contained in the Appendix section. Using Lemma 7 we have that $R_{\Delta u_i \Delta u_i}[t] = 0$ for $t \neq 0$ and $R_{\Delta u_i \Delta u_i}[0]$ is given by

$$R_{y_i y_i}[0] - p^\top \begin{bmatrix} R_{y_i y_i}[0] & R_{y_i y_i}[1] & R_{y_i y_i}[2] \\ R_{y_i y_i}[-1] & R_{y_i y_i}[0] & R_{y_i y_i}[1] \\ R_{y_i y_i}[-2] & R_{y_i y_i}[-1] & R_{y_i y_i}[0] \end{bmatrix} p. \tag{2.13}$$

Measurement Noise

White measurement noise can be represented in the form of (2.6) by setting $C_i[t] = 0$, $D_i[t] = 1$:

$$u_i[t] = y_i[t] + v_i[t]. \tag{2.14}$$

Colored measurement noise with rational spectrum arises when $B_i[t] = 0$, $D_i[t] = 1$, and the matrices $A_i[t]$ and $C_i[t]$ are constant. More generally, the result of causally filtering the signal and then adding noise can be modeled by taking all of the matrices in (2.6) to be constant.

For the corruption model described in (2.14), the perturbation transfer functions are given by:

$$H_i(z) = 1$$
$$\theta_i(z) = \Phi_{v_i v_i}(z).$$

Adversarial Disinformation

This is an example of data-corruption that is pertinent to cyber-security. Here, the true data stream y_i is completely concealed and a new false data stream v_i is introduced. This is an extreme case of (2.6) in which $C_i[t]$ and $D_i[t]$ are zero:

$$u_i[t] = v_i[t] \tag{2.15}$$

Packet Drops

Here the data stream suffers from randomly dropping measurement packets. The corrupted data stream u_i is obtained from y_i as follows:

$$u_i[t] = \begin{cases} y_i[t], & \text{with probability } p_i \\ u_i[t-1], & \text{with probability } (1-p_i) \end{cases} \tag{2.16}$$

Packet drops can be modeled in the form of (2.6) with no noise and matrices given by:

$$\begin{bmatrix} A_i[t] & B_i[t] \\ C_i[t] & D_i[t] \end{bmatrix} = \begin{cases} \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} & \text{with probability } p_i \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} & \text{with probability } 1-p_i. \end{cases} \tag{2.17}$$

The generalized Lyapunov equation becomes:

$$P = p_i P \cdot 0 + (1-p_i) P \cdot 1 + Q \tag{2.18}$$

which has the solution $P = Q/p_i$. Thus, the conditions for Theorem 2 hold, and so u_i is wide-sense stationary. In this case

$$\begin{bmatrix} \bar{A}_i & \bar{B}_i \\ \bar{C}_i & \bar{D}_i \end{bmatrix} = \begin{bmatrix} 1-p_i & p_i \\ 1-p_i & p_i \end{bmatrix} \tag{2.19}$$

so that $H_i(z) = \frac{p_i(1-p_i)}{z-(1-p_i)} + p_i = \frac{p_i}{1-z^{-1}(1-p_i)}$.

The formal description to compute the expression for $\theta_i(z)$ is discussed in Lemma 7 contained in the Appendix section. The application of methods described in the Appendix to derive an expression for $\theta_i(z)$ is cumbersome. However, $\theta_i(z)$ can be calculated directly. Indeed, direct calculation shows that

$$(h_i \star R_{yy} \star h_i^*)[t] = \sum_{j=-\infty}^{|t|} \sum_{k=j}^{\infty} p_i^2 (1-p_i)^{|t|+k-2j} R_{yy}[k] \tag{2.20}$$

while inductive application of (2.16) shows that

$$R_{uu}[t] = (1-p_i)^{|t|} R_{yy}[0] + \sum_{j=1}^{|t|} \sum_{k=j}^{\infty} p_i^2 (1-p_i)^{|t|+k-2j} R_{yy}[k]. \tag{2.21}$$

Here, the sum is interpreted as 0 when $|t| = 0$.

Subtracting (2.20) from (2.21) and taking Z-transforms gives

$$\theta_i(z) = \frac{(1-p_i)^2}{(1-z^{-1}(1-p_i))(1-z(1-p_i))} \cdot \left(R_{yy}[0] + \sum_{j=-\infty}^{0} \sum_{k=j}^{\infty} p_i^2 (1-p_i)^{k-2j} R_{yy}[k] \right). \tag{2.22}$$

2.4 Spurious Links for Perturbed LTI systems

The results reviewed from [74] imply that kin relationships could be inferred from the power spectra of ideal measurements. However, the result of Theorem 2 implies that common types of data corruption cause perturbations to the power spectrum of the observations. In this section, we will show how use of the method from [74] on corrupted data streams leads to the inference of spurious links. In Subsection 2.4.1 we show how spurious links arise in a simple example. Then in Subsection 2.4.2, we characterize the pattern of spurious links that could arise due to data corruption.

(a) Perfect Measurements (b) Unreliable Measurements

Figure 2.2: When node 2 has corrupt measurements an external observer might wrongly infer that the third node is directly influenced by node 1.

2.4.1 Example: Spurious Links due to Data Corruption

Before presenting the general results, an example will be described. Consider the generative graph of a directed chain in Figure 2.2(a). Suppose the measured data-streams are denoted by u_i for node i where $u_i = y_i$ for $i = 1, 3, 4$ (thus no data uncertainty at nodes 1, 3 and 4) and u_2 is related to y_2 via the randomized delay model described in (2.11). In this case, the processes u_i are jointly WSS and the PSD of the vector process $u = (u_1, \cdots, u_4)^\top$ is related to the PSD of the vector process y via:

$$\Phi_{uu}(z) = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & h_2(z) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{H(z)} \Phi_{yy}(z) \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & h_2(z^{-1}) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{H^*(z)} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \theta_2(z) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{D},$$

where h_2 and θ_2 were described in Subsection 2.3.

Note that $\mathcal{D} = b_2 \theta_2 b_2^T$, where $b_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}^T$. Set $\Psi(z) = H(z)\Phi_{yy}(z)H^*(z)$. It follows from the Woodbury matrix identity [137] that

$$\Phi_{uu}^{-1}(z) = \Psi^{-1}(z) - \Psi^{-1}(z) b_2 b_2^T \Psi^{-1}(z) \Delta^{-1}, \tag{2.23}$$

where $\Delta = \theta_2^{-1} + b_2^T \Psi^{-1}(z) b_2$ is a scalar.

Corollary 1 implies that the sparsity pattern of $\Phi_{yy}^{-1}(z)$ is given by:

$$\Phi_{yy}^{-1}(z) = \begin{bmatrix} * & * & 0 & 0 \\ * & * & * & 0 \\ 0 & * & * & * \\ 0 & 0 & * & * \end{bmatrix} \tag{2.24}$$

where $*$ indicates a potential non-zero entry.

Since $H(z)$ is diagonal, it follows that $\Psi^{-1}(z)$ and $\Phi_{yy}^{-1}(z)$ have the same sparsity pattern. Thus, the sparsity pattern of $\Psi^{-1}(z)b_2$ and $\Psi^{-1}(z)b_2b_2^T\Psi^{-1}(z)$ are given by:

$$\Psi^{-1}(z)b_2 = \begin{bmatrix} * \\ * \\ * \\ 0 \end{bmatrix}, \quad \Psi^{-1}(z)b_2b_2^T\Psi^{-1}(z) = \begin{bmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.25}$$

Combining (2.23)-(2.25), it follows that the $\Phi_{uu}^{-1}(z)$ has sparsity pattern given by:

$$\Phi_{uu}^{-1}(z) = \begin{bmatrix} * & * & \boxed{*} & 0 \\ * & * & * & 0 \\ \boxed{*} & * & * & * \\ 0 & 0 & * & * \end{bmatrix}.$$

The extra filled spot in the inverse power spectral density corresponds to a spurious link $1-3$. See Fig. 2.2. Ideally, when there are no measurement uncertainties, measured states of nodes 2 and 4 are sufficient to predict the states of node 3. When node 2 is corrupt, its measurements are unreliable and are not sufficient (together with node 4 measurements) to predict states of 3. In other words, this loss of information will not be able to explain all correlations among the observed nodes. However, node 1 measurements can yield information on node 2. Hence, node 1 measurements together with the noisily measured node 2, and node 4 measurements can yield useful information to predict states of node 3. Similar intuition holds for reasoning why node 3 measurements are useful for predicting states of node 1 and the presence of link $1-3$. However, given uncorrupted measurements of node 3 together with node 2 measurements, node 4 measurements do not contain any useful information to predict states of node 1. This is the intuition behind the absence of link $\{1,4\}$.

2.4.2 Determining Generative Topology from Corrupted Data Streams

In this subsection, we will generalize the insights from the preceding subsection to arbitrary DIMs. The following definitions are needed for the development to follow.

Definition 6 (Path and Intermediate nodes)**.** *Nodes $v_1, v_2, \ldots, v_k \in V$ forms a* path *from v_1 to v_k in an undirected graph $G = (V, A)$ if for every $i = 1, 2, \ldots, k-1$ we have $v_i - v_{i+1}$. The nodes $v_2, v_3, \ldots, v_{k-1}$ are called the* intermediate nodes *in the path.*

Definition 7 (Neighbors $\mathcal{N}$)**.** *Let $G = (V, A)$ be an undirected graph. The neighbor set of node i is given by $\mathcal{N} = \{j : i - j \in A\} \cup \{i\}$.*

Definition 8 (Erroneous Links)**.** *Let $G = (V, A)$ be an undirected graph. An edge or arc $i - j$ is called an erroneous link when it does not belong to A where $i, j \in V$.*

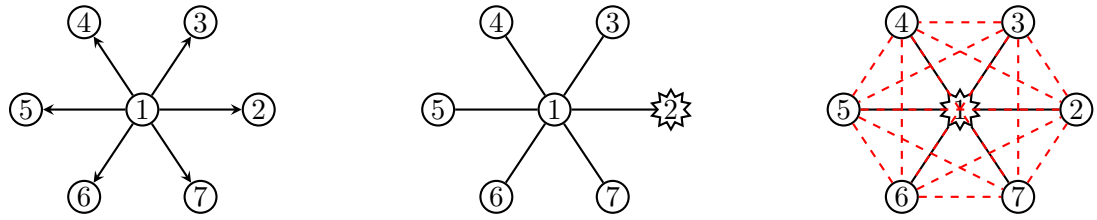
Definition 9 (Perturbed Graph)**.** *Let $G = (V, A)$ be an undirected graph. Suppose $Z \subset V$ is the set of perturbed nodes. Then the perturbed graph of G with respect to set Z is the graph $G_Z = (V, A_Z)$ such that $i - j \in A_Z$ if either $i - j \in A$ or there is a path from i to j in G such that all intermediate nodes are in Z.*

Note that if $Z \subset \hat{Z}$, then $A_Z \subset A_{\hat{Z}}$.

The following theorem is the main result for LTI identification.

Theorem 3. *Consider a DIM $(\mathcal{G}, e)$ consisting of N nodes with the moral graph $G^M = (V, A^M)$. Let $Z = \{v_1, v_2, \ldots, v_n\}$ be the set of n perturbed nodes where each perturbation satisfies* (2.10)*. Then $(\Phi_{uu}^{-1}(z))_{pq} \neq 0$ implies that p and q are neighbors in the perturbed graph G_Z^M.*

Remark 4. *Similar to Remark 3, cases where i and j are kins in the original moral graph, G^M, but $\Phi_{uu}^{-1}(i, j)$ is zero are pathological. $\Phi_{uu}^{-1}(i, j)$ is expressed by terms in $\Phi_{yy}^{-1}, H_l(z)$ and $\theta_l(z)$ where l is a perturbed node. As remarked earlier, the entries in $\mathcal{G}(z)$ and the corruption model described in* (2.6) *must belong to a set of measure zero on space of system parameters such that $\Phi_{uu}^{-1}(i, j)$ is zero. Therefore, except for these restrictive cases, the result in Theorem 3 implies that we can identify the perturbed kin graph.*



(a) Broadcast Architecture

(b) Corrupted leaf

(c) Corrupted Hub

Figure 2.3: This figure shows an extreme example of the effect of data corruption of a single node. 2.3(a) shows the original directed graph. 2.3(b) shows that even if the leaf is corrupted there are no erroneous links introduced. But if the hub is corrupted as shown in 2.3(c) then all the nodes become spuriously correlated.

2.5 Results

Power spectrum estimates were computed after 10^4 simulation time steps. The estimated spectra were then averaged over 100 trials. The red boxes indicate the erroneous links introduced as a result of the network perturbation in addition to the the links in the true moral graph as indicated by the black boxes. For both the networks, the sequences e_i are zero mean white Gaussian noise.

2.5.1 Star Topology

The transfer function for each link is z^{-1}.

Corrupted Leaf

The perturbation considered here is the random delay model, (2.11), on node 2:

$$d_2[t] = \begin{cases} 3, & \text{with probability } 0.65 \\ 1, & \text{with probability } 0.35. \end{cases}$$

$\Phi_{uu}^{-1}(z) =$

$$\begin{bmatrix} 15.02 & \boxed{0.14} & \boxed{1.49} & \boxed{1.49} & \boxed{1.50} & \boxed{1.50} & \boxed{1.45} \\ 0.14 & 1.74 & 0.05 & 0.05 & 0.05 & 0.05 & 0.04 \\ 1.49 & 0.05 & 2.36 & 0.05 & 0.06 & 0.06 & 0.06 \\ 1.49 & 0.05 & 0.05 & 2.35 & 0.06 & 0.05 & 0.06 \\ 1.50 & 0.05 & 0.06 & 0.06 & 2.36 & 0.05 & 0.05 \\ 1.50 & 0.05 & 0.06 & 0.05 & 0.05 & 2.36 & 0.05 \\ 1.45 & 0.04 & 0.06 & 0.06 & 0.05 & 0.05 & 2.34 \end{bmatrix}$$

As predicted by Theorem 3, perturbation of Node 2 for this architecture does not introduce any erroneous links. See Figure 2.3(b).

Corrupted Hub

The perturbation considered here is a random delay on the hub node:

$$d_1[t] = \begin{cases} 2, & \text{with probability } 0.75 \\ 4, & \text{with probability } 0.25. \end{cases}$$

Theorem 3 predicts that perturbing the central node could introduce erroneous links between all of the nodes. See Figure 2.3(c).

$\Phi_{uu}^{-1}(z) =$

$$\begin{bmatrix} 5.08 & \boxed{0.40} & \boxed{0.40} & \boxed{0.40} & \boxed{0.39} & \boxed{0.39} & \boxed{0.38} \\ 0.40 & 2.07 & \boxed{0.27} & \boxed{0.27} & \boxed{0.27} & \boxed{0.26} & \boxed{0.27} \\ 0.40 & 0.27 & 2.08 & \boxed{0.27} & \boxed{0.27} & \boxed{0.28} & \boxed{0.27} \\ 0.40 & 0.27 & 0.27 & 2.07 & \boxed{0.27} & \boxed{0.27} & \boxed{0.27} \\ 0.39 & 0.27 & 0.27 & 0.27 & 2.07 & \boxed{0.27} & \boxed{0.27} \\ 0.39 & 0.26 & 0.28 & 0.27 & 0.27 & 2.08 & \boxed{0.27} \\ 0.38 & 0.27 & 0.27 & 0.27 & 0.27 & 0.27 & 2.08 \end{bmatrix}$$

The estimated power spectrum indicates that indeed, spurious links may be predicted from examination of the power spectrum Φ_{uu}. Therefore, to obtain accurate topology inferences, high fidelity measurements at central nodes would be beneficial.

2.5.2 Chain Topology

The chain topology in Figure 2.4 is considered. The transfer functions are: between nodes 1 and 2, $1.2 + 0.9z^{-1}$, between nodes 2 and 3, $1 + 0.2z^{-1}$, between nodes 3 and 4, $1 - 0.9z^{-1} + 0.3z^{-2}$ and then for the last link z^{-1}. Figure 2.4 In the simulations, nodes

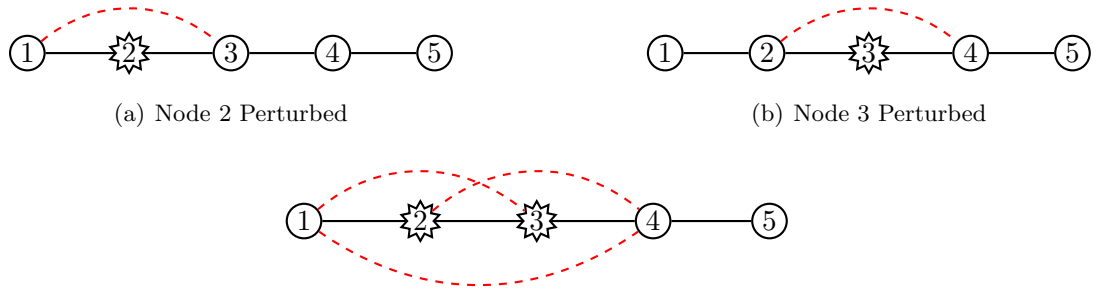
2 and 3 are simultaneously corrupted with the random delay models

$$d_2[t] = \begin{cases} 1, & \text{with probability } 0.83 \\ 2, & \text{with probability } 0.17. \end{cases}$$

$$d_3[t] = \begin{cases} 2, & \text{with probability } 0.85 \\ 4, & \text{with probability } 0.15. \end{cases}$$

$$\Phi_{uu}^{-1}(z) = \begin{bmatrix} 4.23 & \boxed{0.54} & \boxed{0.12} & \boxed{0.25} & 0.05 \\ 0.54 & 1.20 & \boxed{0.16} & \boxed{0.13} & 0.02 \\ 0.12 & 0.16 & 1.06 & \boxed{0.12} & 0.02 \\ 0.25 & 0.13 & 0.12 & 2.22 & \boxed{0.90} \\ 0.05 & 0.02 & 0.02 & 0.90 & 1.42 \end{bmatrix}$$

Perturbation of 2 adds a false relationship between 1 and 3. In addition, perturbation of 3 introduces erroneous relations between the nodes 1 and 4 as well as between 2 and 4. Thus the erroneous relationships could arise between any nodes that are kins of 3 including the already introduced false kins of 3. Despite this cascaded effect the erroneous links remain local in the sense that the dependency of 5 is unaffected.



(a) Node 2 Perturbed

(b) Node 3 Perturbed

(c) Nodes 2 and 3 Perturbed

Figure 2.4: This figure shows how multiple perturbations can lead to a cascade effect as predicted by Theorem 3. Here the original moral graph is a chain. 2.4(a) and 2.4(b) show the erroneous edges that can arise from perturbing a single node. If nodes 2 and 3 are both perturbed, then another erroneous link between 1 and 4 must be added.

2.5.3 Discussion on Robust Network Synthesis

We saw how measurement disturbances can produce errors in the reconstructed network topology. In particular, the results suggest that the effect of data corruption is more drastic for some nodes (like the corrupted hub) than for others. Our results on uncertainty propagation during network reconstruction can be applied to guide strategic placement of high-fidelity sensors and network synthesis that are resilient to and minimize the effects of data corruption. This is significant during planning stage and deployment of monitoring devices in applications such as power distribution [94], water distribution [95], and structural health monitoring [96, 97], where network topology is first identified using sensor measurements for detecting faults and outages. By strategically placing high fidelity sensors, we can mitigate effects of data corruption. For example, in the case of the broadcast architecture, Figure 2.3, placing an ideal sensor at the hub could localize all perturbations.

2.6 Summary

We studied the problem of inferring the network structure of interacting agents from corrupt data-streams. We described general model of data-corruption that introduces an additive term in the power spectra and captures a wide class of measurement uncertainties. We then studied inferring topology of a network of LTI systems from corrupt data-streams. We established that network topology reconstruction from corrupt data streams can result in erroneous links between the nodes. Particularly, we provided exact characterization by proving that the erroneous links are localized to the neighborhood of the perturbed node. Our results show that data corruption gives rise to the appearance of cliques that are localized around the corrupt nodes. The results can be utilized to understand what part of the reconstruction can be trusted and to allocate sensor resources in order to minimize the effects of data corruption.

Chapter 3

Spurious Correlations in Perturbed Markov Random Fields

3.1 Introduction

In the previous chapter, we have shown how perturbing time-series data can give rise to spurious inferences. The analysis focused on network identification for linear time-invariant systems. In this chapter, we will show that spurious links arising from data corruption is a more general phenomenon. Specifically, we will show that the exact same patterns of spurious links from Theorem 3 will arise in a general class of probabilistic graphical models known as *Markov random fields*.

Markov random fields (MRFs) can model a variety of distributions, including continuous and discrete variables. However, our presentation here is restricted to finite-dimensional random variables with well defined probability mass or density functions. Thus, while the class is broad, it does not subsume the analysis from Chapter 2, which deals with infinite-dimensional time-series data.

3.1.1 Related Work

Markov random fields framework is prevalent across several application domains that include gene regulatory networks [10,138], social networks [139], statistical physics [140] image processing [98, 141, 142], robotics [99, 143], finance [144], and climate science [3, 145, 146].

Network structure inference for generic Markov random fields is extensively studied [36, 37, 101, 102, 147–157]. They assume that the observed data are clean without any corruption. No analysis and guarantees of reconstruction are provided when the samples are corrupted.

Robust estimation of Markov random fields are consdiered in [161–165, 168, 169], where a fraction of observations are corrupted. Assuming the knowledge of the fraction of samples being corrupted, sample complexity guarantees to estimate the true structure with confidence are provided. Similarly, assuming the knowledge of statistics of noise model corrupting the observations, sample complexity guarantees to estimate the true structure with confidence are established in [158–160, 166, 167]. Moreover, the analysis of all these works were restricted to either binary valued or Gaussian distributions. We make no assumptions on the probability distributions of the observed data nor do we assume the knowledge of statistics of data corruption model.

Recently, independent of our results, focusing on tree structured Markov random fields, authors in [170–173] analyze reconstructing network structure from noisy observations. They are the closest to our work considering estimation of graphical models using corrupt data without assuming any knowledge of statistics of data corruption model. However, they consider a special case of our problem setup restricting to tree structured models and consider only the neighborhood of leaf nodes to be corrupt. We make no restrictions on the graphical structure and allow multiple corruptions to be located anywhere in the network.

In summary, though network inference of MRF is extensively studied, the effects of data corruption on structure inference are not adequately addressed. Our methods and analysis does not require any structural assumptions on the underlying graph nor any restriction on the distribution of observations nor any knowledge on the noise models. We establish the extent of uncertainty propagation of spurious links during structure inference for generic MRF. We first provide a background on MRF in the following

Section 3.2. Then, exact characterization on the effects of corrupted observations on structure inference is presented in Section 3.3 .

3.2 Description of Markov Random Fields

We provide a formal description of Markov Random Fields and discuss how conditional dependencies among a collection of random variables are encoded by undirected graphs. Our presentation of Markov random fields and probabilistic dependencies will be closely related to graph cliques:

Definition 10 (Clique)**.** *Given an undirected graph $G = (V, A)$, a* clique *is a complete sub-graph formed by a set of vertices $b \subset V$ such that for all distinct $i, j \in b$ there exists $i - j \in A$.*

As an example of a Markov random field, consider a finite-dimensional version of the model from (2.4):

$$\begin{aligned} y_1 &= e_1 \\ y_2 &= M_{21}y_1 + e_2 \\ y_3 &= M_{31}y_1 + e_3 \\ y_4 &= M_{42}y_2 + M_{43}y_3 + e_4 \\ y_5 &= M_{54}y_4 + e_5 \end{aligned} \tag{3.1}$$

with

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ M_{21} & 0 & 0 & 0 & 0 \\ M_{31} & 0 & 0 & 0 & 0 \\ 0 & M_{42} & M_{43} & 0 & 0 \\ 0 & 0 & 0 & M_{54} & 0 \end{bmatrix}.$$

Here, we take e_i to be independent Gaussian vectors with mean 0 and covariance E_i. When only a finite amount of time series data has been collected for the system in (2.4), the relationship between the data points can be modeled as in (3.1).

Now we will see how the structure of the probabilistic relationships between the variables, y_i are encoded in the corresponding moral graph from Fig. 2.1(b). If $y = \begin{bmatrix} y_1 & \cdots & y_5 \end{bmatrix}^\top$ and $e = \begin{bmatrix} e_1 & \cdots & e_5 \end{bmatrix}^\top$, then $y = (I - M)^{-1}e$. Use the notation

$\|x\|^2_{E_i^{-1}} = x^\top E_i^{-1} x$. Then direct calculation shows that the density of y factorizes as

$$p(y) = c \cdot \exp\left(-\frac{1}{2}\|y_1\|^2_{E_1^{-1}} - \frac{1}{2}\|y_2 - M_{21}y_1\|^2_{E_2^{-1}} \cdot -\frac{1}{2}\|y_3 - M_{31}y_1\|^2_{E_3^{-1}}\right) \cdot \exp\left(-\frac{1}{2}\|y_4 - M_{42}y_2 - M_{43}y_3\|^2_{E_4^{-1}}\right) \cdot \exp\left((-\frac{1}{2}\|y_5 - M_{54}y_4\|^2_{E_5^{-1}}\right). \quad (3.2)$$

Note that the exponential factors contain variables $\{y_1, y_2, y_3\}$, $\{y_2, y_3, y_4\}$, and $\{y_4, y_5\}$. These variable groupings correspond precisely to the maximal cliques in the moral graph from Fig. 2.1(b).

As we will discuss below, having a distribution that factorizes with respect to a graph is a sufficient condition for being a Markov random field. See also [174]. A generalization of the construction of (3.1) shows that finite collections of time-series data can always be viewed as Markov random fields.

To formally define Markov random fields, we need some extra notation and terminology. Let Y be a collection of variables, $Y = \{y_1, \ldots, y_{|V|}\}$ corresponding to nodes of a graph, $G = (V, A)$. If $S \subset V$, then we use the notation $Y_S = \{y_i | i \in S\}$.

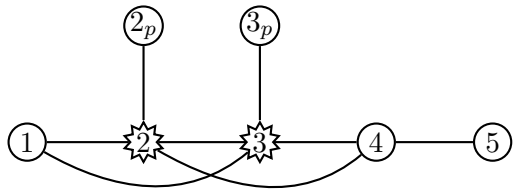


Figure 3.1: Markov random field G^J with perturbed nodes.

Definition 11 (Separation)**.** *Suppose $G = (V, A)$ is an undirected graph. Suppose, a, b, c are disjoint subsets of V. Then, a and b are* separated *given c if all paths from a to b must pass through c.*

When a and b are separated given c, we write $\text{sep}(a, b \mid c)$.

Definition 12 (Markov random fields)**.** *Let Y be a collection of random variables associated with the nodes of an undirected graph, $G = (V, A)$. The variables Y are called a*

Markov random field *with respect to G if Y_a and Y_b are conditionally independent given Y_c whenever $sep(a, b \mid c)$ holds.*

A useful sufficient condition for Y to be a Markov random field with respect to G is for the distribution to factorize into terms corresponding to cliques in the graph. This condition was used in the example above. See [174] for more details.

Definition 13 (Clique Factorization)**.** *Suppose that Q is a collection of subsets of V such that each $q \in Q$ forms clique in G. Let $P(Y)$ denote the joint probability distribution of the random variables Y. Then, we say Y* factorizes *according to G, if for every $q \in Q$, there exists non-negative functions Ψ_q that are functions of random variables in q such that,*

$$P(Y) = \prod_{q \in Q} \Psi_q(Y_q) \tag{3.3}$$

3.3 Inferring Erroneous Links

Now we will describe the effects of data-corruption on inferring the undirected graph structure from measured data. In our work on time-series models, we assumed that individual data streams were perturbed independently. Here we will define a natural analog of independent perturbations for Markov random fields. However, the perturbation models could be non-linear.

Let Y be a Markov random field that factorizes with respect to a graph $G = (V, A)$. Let $Z \subset V$ be the set of perturbed nodes. For each perturbed node, $i \in Z$, we draw a new node i_p, draw an edge $i - i_p$, and denote the corresponding perturbed variable by u_i. Note that for all $i \in Z$ we only observe the perturbed version, u_i and not the original variable y_i. The probabilistic relationships between the original variable, y_i, and the perturbed variable, u_i, is given by $\Psi_{ii_p}(y_i, u_i) \geq 0$. Let $Z_p = \{i_p : i \in Z\}$ and let U_Z denote the set of perturbed variables. Then the joint distribution between Y and U_Z can be described as:

$$P(Y, U_Z) = \prod_{q \in Q} \Psi_q(Y_q) \cdot \prod_{i \in Z} \Psi_{ii_p}(y_i, u_i). \tag{3.4}$$

Since the node pairs, $\{i, i_p\}$ are cliques, the construction above shows that the joint variables (Y, U_Z) form a Markov random field with respect to $G^J = (V \cup Z_p, A \cup \{i - i_p : \forall i \in Z\})$. See figure 3.1.

Due to data corruption, only the variables $Y_{\bar{Z}}$ and U_Z are observed, where $\bar{Z} = V \setminus Z$. The next lemma shows that $(Y_{\bar{Z}}, U_Z)$ is also a Markov random field, with graph described by the perturbed graph.

Lemma 1. *Let Y be a Markov random field with respect to $G = (V, A)$. Let $Z \subset V$ be a set of perturbed nodes and let $\bar{Z} = V \setminus Z$ be the unperturbed nodes. Assume that the joint distribution of Y and the perturbed variables U_Z factorizes as in (3.4). Then the collection of observed variables $(Y_{\bar{Z}}, U_Z)$ factorizes with respect to the perturbed graph G_Z from Definition 9.*

In our model, we have assumed that the variables corresponding to corrupted nodes, y_i for $i \in Z$, are hidden. Then Lemma 1 shows that marginalizing out the variables y_i introduces new probabilistic relationships between the neighbors of y_i. The new links between variables are precisely described by the perturbed graph construction of Theorem 3. Note that *any* method that attempts to reconstruct the graphical structure of the Markov random field based only on the observed data that contains corrupt data will be likely to detect spurious relationships.

Below, we will show that if $P(Y_{\bar{Z}}, U_Z)$ is positive everywhere, then the perturbed graph exactly characterizes the conditional independence of the observed nodes. To present this strengthened version of Lemma 1, some definitions are required.

Definition 14 (Pairwise Markov property)**.** *Suppose $G = (V, A)$ is an undirected graph whose N nodes represent random variables $y_1, \ldots, y_N$. Let $Y = \{y_1, \ldots, y_N\}$.* Pairwise Markov property *associated with G holds, if for any non-adjacent vertices i, j, we have that $sep(i, j | V \setminus \{i, j\})$ implies that y_i and y_j are conditionally independent given $Y \setminus \{y_i, y_j\}$.*

As in the discussion of LTI systems, it is convenient to identify the observed but unperturbed variables $Y_{\bar{Z}}$ with $U_{\bar{Z}}$ so that the collection of observed variables can be denoted by $U = (U_{\bar{Z}}, U_Z)$. The following theorem characterizes the effects of data corruption in inferring structure of Markov random fields.

Theorem 4. *Let Y be a set of random variables that factorize according to graph $G = (V, A)$. Suppose, $Z \subset V$, is a set of perturbed nodes such that the joint distribution (Y, U_Z) factorizes as in (3.4). Let U denote the set of all observed variables and assume*

that $P(U)$ is positive everywhere. Define $U_{\overline{ij}} := U \setminus \{u_i, u_j\}$. Then, $i - j$ is an edge in the perturbed graph, G_Z, if and only if u_i is not conditionally independent of u_j given $U_{\overline{ij}}$.

The above result concludes that we infer spurious correlations among the observed data when there is data corruption. As only unreliable data are available, the states corresponding to ideal measurements are hidden and marginalized. This introduces spurious probabilistic relationships among the neighbors of the corrupt nodes and thus, these spurious correlations inferred are localized to the neighborhood of the corrupt node.

3.4 Summary

We studied the influence of data corruption on Markov random field models. Here, we found that our characterization of erroneous links for LTI systems from previous chapter precisely characterized the spurious relationships that can arise in Markov random fields. Thus, we showed that spurious links arising from data corruption is a more general phenomenon.

Chapter 4

Effects of Data Corruption on Network Identification for Nonlinear Systems

4.1 Introduction

In this chapter, we switch our attention to networks of nonlinear dynamical interactions and causal inference methods. In particular, we analyze how spurious probabilistic connections are inferred in networks of nonlinear dynamical interactions due to data-corruption. Thus, we will further reinforce that spurious links arising from data corruption is a more general phenomenon. We begin by presenting an overview of related work in network identification for nonlinear dynamics and causal inference methods.

4.1.1 Related Work

Authors in [74] leveraged multivariate Wiener filters to reconstruct the undirected graph of the generative network model. Moreover, assuming that the interaction dynamics are *strictly causal* and using multivariate estimation based on a Granger filter, it was shown that the directed interaction structure can be accurately recovered. However, results assume data to be uncorrupted with the interaction between agents governed via linear time-invariant (LTI) dynamics.

For a network of interacting agents with nonlinear dynamic dependencies and strictly causal interactions, the authors in [78] proposed the use of directed information to determine the directed structure of the network. Sufficient conditions to recover the directed structure are provided. Recently, [175], [176], [177] defined and used *information transfer* to determine underlying causal interactions in dynamical systems. However, it is assumed that the data-streams are ideal with no distortions. [76] and [136] identify causal dependencies in network of LTI systems driven by unknown intrinsic noise inputs. However, in this work we consider nonlinear dependencies and study the problem of network reconstruction from corrupt data-streams.

The authors in [133], [134] use dynamical structure functions (DSF) for network recosntruction [135] and consider measurement noise and non-linearities in the network dynamics. The proposed method first finds optimal DSF for all possible Boolean structures and then adopt a model selection procedure to determine the best estimate. The authors concluded that the performance of their algorithms degrades as noise, network size and non-linearities increase. However, a precise characterization of drawing spurious inferences in structure is not provided. In this work, we provide exact location of spurious links that arise during directed information-based network reconstruction from corrupt data-streams.

Despite its significance, little is known on the effects of measurement uncertainties on network identification. Assuming that the network structure is known, *errors-in-variables* framework for system identification with additive sensor noise is studied in [129], [104]. However, in this work we do not assume that the interaction structure is known. Recently in [178], the issues of observation noise and undersampling on causal discovery from time-series data has been addressed. Although authors concluded that spurious links can be inferred, a rigorous characterization of such links was not proven nor a generalization of corruption models was provided. In [179] focusing on networks with LTI interactions, authors provided characterization of the extent of spurious links that can appear due to data-corruption. However, the analysis is restricted to LTI systems and determine the undirected structure of the networked system, and not to deduce the directions.

4.1.2 Our Contribution

In this work, we focus on identifying the Boolean structure of the network using non-invasive or passive means from corrupt data-streams. We consider networks with nonlinear and strictly causal dynamical interactions. Moreover, the endogenous noise exciting the system are not measured and hence, we assume a *blind* approach [85]. General analysis of network structure employing passive and and blind means with nonlinearities is challenging. We make an assumption that the endogenous noise affecting one node is independent of another and thus we deal with *target specified* network reconstruction.

We provide necessary and sufficient conditions that delineates the effects of data corruption on the directed network structure inferred using directed information. We present a tight characterization for the spurious links that arise due to corruption of data-streams by determining their location and orientation. Often, the knowledge of influence structure is required a priori to perform system identification in networked systems [103], [104], [114]. Thus, our results serve as a necessary first step in understanding what part of network reconstruction can be trusted to facilitate accurate system identification.

In [180], preliminary results that characterized the spurious links, in the framework of this work are provided. However, the analysis was limited to dynamical interactions such that every node was dependent dynamically on the entire history (strict) of its *parent* nodes. In this chapter, we consider a general class of non-linear systems by relaxing the above assumption on dynamics. Moreover, we provide detailed and rigorous proofs to generalize the results obtained in [180] wherein only a proof sketch was provided. In addition, we establish convergence results for the estimator that we use to determine conditional directed information.

We review needed graph theory notions and describe the framework for generative models in Section 4.2. In Section 4.3, we provide models to characterize corruption of data-streams. Directed network structure inference methods and effects of data corruption are described in Section 4.4.

Notations

Upper case letter Y denotes a random variable (r.v) while lower case letter y denotes a realization of r.v Y.
Caligraphic letter $\mathcal{Y}$ denotes the alphabet of r.v Y.
$y[\cdot]$ denotes a sequence and $y^{(t)}$ denotes the sequence $y[0], y[1], \dots y[t]$.
P_X represents the probability mass function (PMF) of a discrete random variable X or denotes the probability density function (PDF) of a continuous random variable X.
$i - j$ denotes one of $i \to j$ or $j \to i$.

4.2 Preliminaries

4.2.1 Graph Theory Definitions

This subsection gives a list of standard terminology from graphical models. It can be used as a reference for following sections. For further details, see [36].

Definition 15 (Trail/Path)**.** *Nodes $v_1, v_2, \dots, v_k \in V$ forms a* trail *or a* path *in a directed graph, G, if for every $i = 1, 2, \dots, k-1$ we have $v_i - v_{i+1}$.*

Definition 16 (Chain)**.** *In a directed graph G, a* chain *from node v_i to node v_j comprises of a sequence of k nodes such that $v_i \to W_1 \to \cdots \to W_{k-2} \to v_j$ holds in G.*

Definition 17 (Descendants and Ancestors)**.** *Suppose there exists a chain from a node v_j to v_k in a directed graph, G. Then, v_k is called a* descendant *of node v_j and v_j is called an* ancestor *of v_k.*

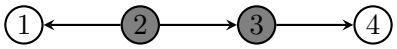
Definition 18 (Collider)**.** *A node v_k is a* collider *in a directed graph, G, if there are two other nodes v_i, v_j such that $v_i \to v_k \leftarrow v_j$ holds.*

Definition 19 (Active Trail)**.** *In a directed graph G, a trail $v_1 - v_2 - \cdots - v_n$ is* active *given a set of nodes Z if one of the following statements holds for $m \in \{2, \dots, n-1\}$ and every triple $v_{m-1} - v_m - v_{m+1}$ along the trail:*

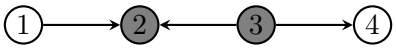
a) If v_m is not a collider, then $v_m \notin Z$.

b) If v_m is a collider, then v_m or one of its descendants is in Z.

See Figure 4.1 for an illustration.



(a) Trail connecting 1 and 4 is active given $Z = \{\}$.



(b) Trail connecting 1 and 4 is active given $Z = \{2\}$.

Figure 4.1: This figure shows when the trail connecting nodes 1 and 4 is active given Z.

Definition 20 (d-separation)**.** *Let X, Y and Z be a set of nodes in a directed graph, G. In G, X and Y are* d-separated *by Z if and only if there is no active trail between any $x \in X$ and any $y \in Y$ given Z. It is denoted as d-sep$(X, Y \mid Z)$.*

Definition 21 (Directed Cycle)**.** *A* directed cycle *from a node v_i to v_i in a directed graph, G, has the form $v_i \to W_1 \to \cdots \to W_k \to v_i$ for some set of nodes $\{W_n\}_{n=1}^k$ in G.*

Definition 22 (Directed Acyclic Graph)**.** *A directed graph with no directed cycles is called a* directed acyclic graph *(DAG).*

Definition 23 (Bayesian Network)**.** *Suppose $G = (V, A)$ is a DAG whose N nodes represent random variables $a_1, \ldots, a_N$. G is called a Bayesian Network (BN) if for any three subsets X, Y and Z of V, d-sep$(X, Y \mid Z)$ implies X is independent of Y given Z.*

Definition 24 (Faithful Bayesian network)**.** *Suppose $G = (V, A)$ is a DAG whose N nodes represent random variables $a_1, \ldots, a_N$. G is called a* Faithful Bayesian network *if for any three subsets X, Y and Z of V, it holds that X and Y are independent given Z, if and only if d-sep$(X, Y \mid Z)$ is true.*

4.2.2 Generative Model

In this subsection, the *generative model* that is assumed to generate the measured data is described. Consider N agents that interact over a network. For each agent i, we associate a discrete time sequence $Y_i[\cdot]$ and a sequence $E_i[\cdot]$. We consider E_i such that

P_{E_i} exists if $E_i[t]$ belongs to a continuous alphabet. The process E_i is considered to be target-specific, that is, E_i is innate to agent i and thus E_i is independent of E_j if $i \neq j$. Moreover, E_i is considered to be uncorrelated across time. Let Y denote the set of all random process $\{Y_1, \ldots, Y_N\}$ with a parent set $\mathcal{P}'(i)$ defined for $i = 1, \ldots, N$. We consider strictly causal nonlinear dynamical relations. The generative model takes the form:

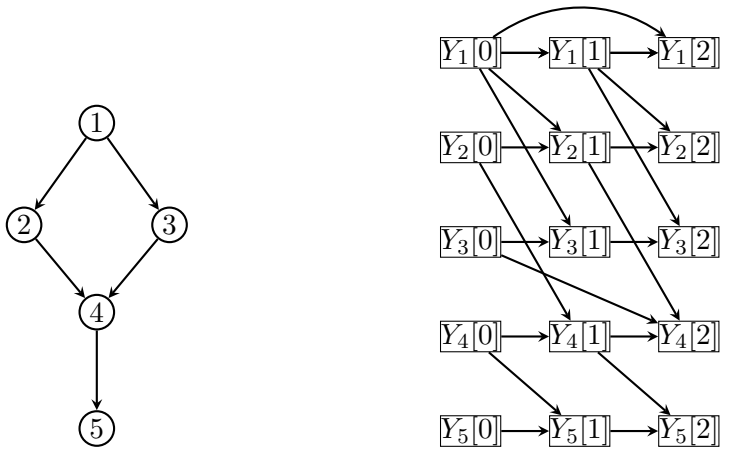
$$Y_i[t] = f_i\left(Y_i^{(t-1)}, \bigcup_{j \in \mathcal{P}'(i)} Y_j^{(t-1)}, E_i[t]\right), \tag{4.1}$$

where f_i's can be any nonlinear function such that P_{Y_i} is well defined if $Y_i[t]$ takes values in a continuous alphabet. f_i is a multivariate function that maps the past measurements of parent nodes of i, $\{Y_j^{(t-1)} : j \in \mathcal{P}'(i)\}$, previous measurements of the node i in $Y_i^{(t-1)}$, and the present realization of process noise, $E_i[t]$, to the present measurement of agent i, $Y_i[t]$.

For an illustration, consider the dynamics of a generative model described by:

$$\begin{aligned}
Y_1[t] &= Y_1[t-1]Y_1[t-2] + E_1[t], \\
Y_2[t] &= Y_1[t-1] \cdot Y_2[t-1] + E_2[t], \\
Y_3[t] &= (Y_1[t-1] + Y_3[t-1]) \cdot E_3[t], \\
Y_4[t] &= Y_2[t-1]^2 + Y_3[t-2] + Y_4[t-1] + E_4[t], \\
Y_5[t] &= Y_5[t-1] \cdot Y_4[t-1] + E_5[t].
\end{aligned} \tag{4.2}$$

We remark that for any time instant t, the parent set $\mathcal{P}'(i)$ is thus not dependent on time.



(a) Generative Graph G

(b) DBN G' for 3 time slices

Figure 4.2: This figure shows (a) generative graph, (b) its associated DBN for 3 time slices.

4.2.3 Graphical Representation

Here we describe how networks of dynamical systems are represented by graphs.

Generative Graph

The structural description of (4.1) induces a *generative graph* $G = (V, A)$ formed by identifying each vertex v_i in V with random process Y_i and the set of directed links, A, obtained by introducing a directed link from every element in the parent set $\mathcal{P}'(i)$ of agent i to i. Note that we do not show $i \to i$ in the generative graph and neither do we show the processes E_i. The generative graph describes the relationships between the stochastic processes in Y.

The generative graph associated with the example described in (4.2) is given by Fig. 4.2(a). When the time variable is unraveled we obtain the Dynamic Bayesian Network.

Dynamic Bayesian Network (DBN)

Let $G = (V, A)$ be a generative graph. Let Y_i be as defined in (4.1) for all $i \in V$. Suppose all discrete time sequences have a finite horizon assumed to be T. Let $S_{ij}[t] = \{t' :$

$Y_j[t'] \in Y_j^{(t-1)}$as an argument of f_i in expression of $Y_i[t]$ in (4.1)} for all $j \in \mathcal{P}'(i) \cup \{i\}$ and for all t. Consider the graph $G' = (V', A')$ where $V' = \left(\bigcup_{\substack{i \in V \\ t \in \{0,1,\ldots T\}}} Y_i[t] \right)$ and

$$A' = \bigcup_{\substack{i \in V \\ t \in \{0,1,\ldots T\}}} \left(\bigcup_{j \in \mathcal{P}'(i) \cup \{i\}} \left(\bigcup_{k \in S_{ij}[t]} Y_j[k] \rightarrow Y_i[t] \right) \right)$$

The joint distribution of $Y^{(T)}$ is given by:

$$P_{Y^{(T)}} = P_{Y_1[0]} \ldots P_{Y_N[0]} \prod_{t=1}^{T} \prod_{i=1}^{N} P_{Y_i[t]|\mathcal{P}(Y_i[t])}, \tag{4.3}$$

where the parents of $Y_i[t]$ are obtained from G'. It can be shown that G' is the Bayesian network for the random variables $\{Y_i[t] : t = 0, 1, 2, \ldots, T,\ i = 1, 2, \ldots, N\}$ and is considered the *Dynamic Bayesian Network* for $\{Y_i : i = 1, 2, \ldots, N\}$(see [36]). Figure 4.2(b) represents the DBN for the system in (4.2) for three time steps.

4.3 Uncertainty Description

In this section we provide a description for how uncertainty affects the time-series Y_i. We interchangeably use corruption or perturbation to denote uncertainties in data-streams.

4.3.1 General Perturbation Models

Consider i^{th} node in a generative graph and it's associated unperturbed time-series Y_i. The corrupt data-stream U_i associated with i follows:

$$U_i[t] = g_i(Y_i^{(t)}, U_i^{(t-1)}, \zeta_i[t]), \tag{4.4}$$

where g_i can be any multivariate function that maps the present and past values of uncorrupted data-streams in $Y_i^{(t)}$, the present value of an independent random process $\zeta_i[t]$, and past corrupt measurements in $U_i^{(t-1)}$ to the current corrupt measurement, $U_i[t]$, such that P_{U_i} exists if $U_i[t]$ takes values in a continuous alphabet. $\zeta_i[t]$ is such that P_{ζ_i} exists if $\zeta_i[t]$ belongs to a continuous alphabet, and is independent of E_i, Y_i for all $i \in 1, \cdots, N$, and $\zeta_j[t]$ for $i \neq j$. Note that the above perturbation model admits nonlinear dependecies and is more general than the corruption models discussed

in Chapter 2. We highlight a few important perturbation models that are practically relevant.

Temporal Uncertainty

Consider a node i in a generative graph. Suppose t is the true clock index but the node i measures a noisy clock index which is given by a random process, $\zeta_i[t]$. One such probabilistic model is given by the following IID Bernoulli process:

$$\zeta_i[t] = \begin{cases} d_1, & \text{with probability } p_i \\ d_2, & \text{with probability } (1-p_i), \end{cases}$$

where d_1 and d_2 are any non-positive integers such that at least one of d_1 and d_2 are not equal to 0. Randomized delays in information transmission can be modeled as a convolution operation with the impulse function $\delta[t]$ shifted by $\zeta_i[t]$ as :

$$U_i[t] = \delta[t + \zeta_i[t]] * Y_i[t], \tag{4.5}$$

where,

$$\delta[t] = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0. \end{cases}$$

Noisy Filtering

Given a node i in a generative graph, the data-stream Y_i is causally filtered and corrupted with independent measurement noise $\zeta_i[\cdot]$. This perturbation model is described by:

$$U_i[t] = (L_i * Y_i)[t] + \zeta_i[t], \tag{4.6}$$

where L_i is a stable causal linear time invariant filter.

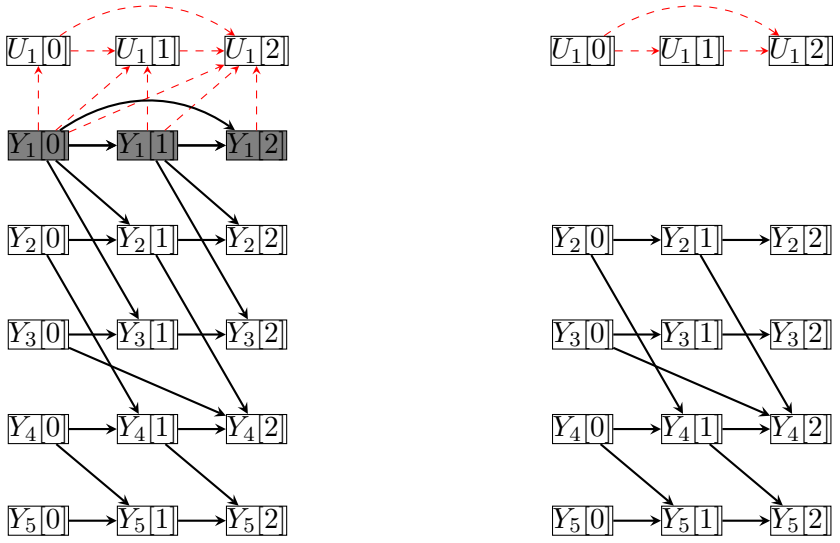
Packet Drops

The measurement $U_i[t]$ corresponding to an ideal data-point $Y_i[t]$ packet reception at time t can be stochastically modeled as:

$$U_i[t] = \begin{cases} Y_i[t], & \text{with probability } p_i \\ U_i[t-1], & \text{with probability } (1-p_i). \end{cases} \tag{4.7}$$

Consider an IID Bernoulli process $\zeta_i[t]$ described by success probability, p_i. The corruption model in (4.4) takes the form:

$$U_i[t] = \zeta_i[t]Y_i[t] + (1 - \zeta_i[t])U_i[t-1]. \tag{4.8}$$



(a) PDBN for 3 time slices

(b) Measured Data-Stream

Figure 4.3: Figure (a) shows Perturbed DBN G'_Z for 3 time slices when node 1 is corrupt. Node 1 ideal stream denoted by Y_1 is shaded because it is not measured. Figure (b) shows only the measured data-streams and their causal relations.

4.3.2 Perturbed Dynamic Bayesian Network

Here, we provide a discussion on how the dynamic Bayesian network associated with the measured data-streams gets altered when the data-streams are subject to corruption. Note that the measured data-streams only includes the corrupted time-series for the nodes that are corrupted and the data-streams for those nodes that are not corrupted. The uncorrupted time-series for the corrupted nodes are not measured. When the time variable is unraveled we obtain the perturbed DBN (PDBN) that depicts the causal dependencies between the true data-streams for the network, and in addition shows the dependencies between the uncorrupted measurements and the corrupted values for the corrupted nodes, and between the corrupted measurements for each corrupted node.

Thus, the perturbed DBN is the union of DBN and the causal dependencies for the time-series associated with the corrupted node. Figure 4.3(a) shows an example of a PDBN corresponding to the generative graph in Fig. 4.2(a) for three time slices. Here, node 1 data-streams are corrupt following a noisy filtering model described in (4.6).

Consider a generative graph $G = (V, A)$. Let Y_i be as defined in (4.1) for all $i \in V$. Suppose all discrete time sequences have a finite horizon assumed to be T. Let $G' = (V', A')$ be the associated dynamic Bayesian network. Suppose $Z \subset V$ is the set of perturbed nodes with perturbation model described in (4.4). For $i \in Z$, the measured (corrupt) data-stream corresponding to agent i, U_i, is related to Y_i via (4.4). Let $U_Z = \{U_i\}_{i\in Z}$ and $Y_{\bar{Z}} = \{Y_j\}_{j\in\bar{Z}}$ where $\bar{Z} = V \backslash Z$. Due to corruption only U_Z and $Y_{\bar{Z}}$ are measured and observed. Denote the measured data-streams by $\mathcal{W} = U_Z \cup Y_{\bar{Z}}$. For all $j \in Z$, let $SU_j[t] = \{t' : U_j[t'] \in U_j^{(t-1)}$ is an argument of g_i in the expression of $U_j[t]$ in (4.4)$\}$ and let $SY_j[t] = \{t' : Y_j[t'] \in Y_j^{(t)}$ is an argument of g_i in the expression of $U_j[t]$ in (4.4)$\}$ for all t. Consider the graph $G'_Z = (V'_Z, A'_Z)$ where $V'_Z = V' \cup \left(\bigcup_{\substack{k\in Z \\ t\in\{0,1,\ldots T\}}} U_k[t] \right)$ and $A'_Z = A' \cup \left(\bigcup_{\substack{k\in Z \\ i\in SY_k[t]}} Y_k[i] \to U_k[t] \right) \cup \left(\bigcup_{\substack{k\in Z \\ i\in SU_k[t]}} U_k[i] \to U_k[t] \right)$ for all $t \in \{0, 1, 2, \ldots, T\}$. Note that the vertex set V'_Z consists of all measurements given by the set $\mathcal{W}$, and the uncorrupted versions Y_k of the corrupted versions U_k for $k \in Z$.

Consider the set of random variables, $R = \{Y_i[t] : i \in \{1, 2, \ldots, N\}$ and $t \in \{0, 1, 2, 3, \ldots, T\}\} \cup \{U_i[t] : i \in \{1, 2, \ldots, N\}$ and $t \in \{0, 1, 2, 3, \ldots, T\}\}$. The joint distribution P_R is given by:

$$P_R = \left(\prod_{i\in V} P_{U_i[0]} \right) \cdot \left(\prod_{j\in Z} P_{Y_j[0]} \right) \cdot \left(\prod_{t=1}^{T} \prod_{i=1}^{N} P_{U_i[t]|\mathcal{P}(U_i[t])} \right) \cdot \left(\prod_{t=1}^{T} \prod_{j=1}^{N} P_{Y_j[t]|\mathcal{P}(Y_j[t])} \right), \quad (4.9)$$

where the parents of $U_i[t], Y_j[t]$ are obtained from G'_Z. G'_Z is the Bayesian Network for the random variables R and is considered as the perturbed DBN associated with $U_Z \cup Y$.

Remark 5. *Above, the discrete time sequences were considered to have finite horizon only to illustrate DBN and PDBN. The main results in this work characterizing the structure inference from corrupt data-streams holds for any horizon.*

4.4 Structure Identification

4.4.1 Structure Inference from Ideal Data-Streams

First, we recall how the structure of a generative graph can be inferred using directed information in the case of ideal data-streams. Consider a generative graph G with N nodes and let Y denote the collection of N data-streams that are measured. The authors in [78] defined and applied causally conditioned directed information (DI) in a network of dynamically interacting agents to determine if a process causally influences another. A slightly modified definition of DI as defined in [78] is:

Definition 25 (Causally Conditioned Directed Information)**.** *The causally conditioned directed information (DI) from data-stream Y_j to Y_i is given by:*

$$I(Y_j \to Y_i \parallel Y_{\overline{ij}}) = \mathbb{E}\left[\log \frac{P_{Y_i \| Y_j, Y_{\overline{ij}}}}{P_{Y_i \| Y_{\overline{ij}}}}\right], \tag{4.10}$$

where $P_{Y_i \| Y_j, Y_{\overline{ij}}} = \prod_{t=1}^{T} P_{Y_i[t] | Y_i^{(t-1)}, Y_j^{(t-1)}, Y_{\overline{ij}}^{(t-1)}}$, $P_{Y_i \| Y_{\overline{ij}}} = \prod_{t=1}^{T} P_{Y_i[t] | Y_i^{(t-1)}, Y_{\overline{ij}}^{(t-1)}}$ *and* $Y_{\overline{ij}} = Y \setminus \{Y_i, Y_j\}$.

For the rest of the chapter, we drop the word 'causally' for convenience. Note that the conditional DI from Y_j to Y_i is positive if and only if the history of Y_j gives information about $Y_i[t]$ that could not have been obtained from Y_i's own history and the other signals from the network. So, if there is no directed edge from $j \to i$ in G, then we have $I(Y_j \to Y_i \parallel Y_{\overline{ij}}) = 0$.

The following theorem was proved in [78] that specifies a necessary and sufficient condition to detect a presence of link in the generative graph.

Theorem 5. *A directed edge from j to i exists in the directed graph G if and only if $I(Y_j \to Y_i \parallel Y_{\overline{ij}}) > 0$.*

Remark 6. *In [78], the authors assume positive distribution for the random processes in Y. The distribution is positive if $P_Y > 0$ for all joint sequences Y. This assumption avoids pathologies that arise in deterministic systems. For example, if $Y_1[t]$ are IID random variables, and $Y_2[t] = Y_1[t-1]$ and $Y_3[t] = Y_2[t-1]$, then $I(Y_2 \to Y_3 \parallel Y_1) = 0$, even though Y_3 depends on Y_2 The positivity assumption ensures that the computed expectations are non-negative and hence avoids false negatives for true edges in the generative graph and are therefore detected.*

4.4.2 Main Result: Inferring Directed Graphs from Corrupt Data-streams

In this section we characterize the spurious edges that arise when using conditional DI to estimate network structure. In particular, we will show that under appropriate hypotheses, the estimated edges precisely correspond to edges in the *perturbed graph*, defined next.

Definition 26 (Perturbed Graph)**.** *Let $G = (V, A)$ be a generative graph. Suppose $Z \subset V$ is the set of perturbed nodes with each perturbation model admitting a description provided by* (4.4)*. The perturbed graph, $G_Z = (V, A_Z)$, is a directed graph where there is an edge $i \to j \in A_Z$ if and only if there is a trail, $trl_G : i = v_1 - v_2 - \cdots - v_{k-1} - v_k = j$ in G such that the following conditions hold:*

P1) If $j \notin Z$, then $v_{k-1} \to j \in A$.

P2) For $m \in \{2, 3, \ldots, k-1\}$, if $v_{m-1} \to v_m \leftarrow v_{m+1}$, and $v_m \notin Z$, then $v_{m+1} \in Z$.

P3) If v_m is a node such that $v_{m-1} - v_m - v_{m+1}$ is a sub-path of the path $v_1 - \ldots - v_k$ and v_m is not a collider, then $v_m \in Z$.

Remark 7. *Note that the existence of a trail that does not meet the 'if' conditions in P1), P2) and P3) guarantees that $i \to j \in A_Z$. For example, if $i \to j \in A$ then $i \to j \in A_Z$. Indeed, if $j \notin Z$ then $i \to j \in A_Z$ by condition P1).Conditions P2) and P3) are not applicable. On the other hand, if $j \in Z$, then none of the conditions P1), P2) or P3) are applicable to the trail $i \to j$. So, $i \to j \in A_Z$.*

Definition 27 (Spurious Links)**.** *Let $G = (V, A)$ be a generative graph, $Z \subset V$ be the set of perturbed nodes and $G_Z = (V, A_Z)$ be the perturbed graph. Spurious links are those links $i \to j \in A_Z$ that do not belong to A.*

The conditions in P1-P3 specifies a path characterization based on the location of corrupt nodes. This defines the paths through which spurious probabilistic relations are introduced due to data corruption. These probabilistic relations are captured by trails in PDBN that become active due to data-corruption. The following theorem precisely gives a relationship between the active trails in PDBN and the directed edges in the perturbed graph.

Theorem 6. *Consider a generative graph, $G = (V, A)$, consisting of N nodes. Let $Z = \{v_1, \ldots, v_n\} \subset V$ be the set of n perturbed nodes where each perturbation is described by (4.4). Denote the data-streams as follows: $U_Z := \{U_i\}_{i \in Z}$ and $Y_{\bar{Z}} := \{Y_j\}_{j \in \bar{Z}}$ where $\bar{Z} = V \backslash Z$. Let the measured data-streams be $\mathcal{W} = U_Z \cup Y_{\bar{Z}} = \{W_1, W_2, \ldots, W_N\}$. Let the perturbed graph be $G_Z = (V, A_Z)$ and its associated perturbed DBN be $G'_Z = (V'_Z, A'_Z)$. If $i \to j \notin A_Z$, then d-sep$(W_j[t], W_i^{(t-1)} \mid \{W_i^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\})$ holds in G'_Z for all $t > 0$.*

Proof. The proof is given in appendix C.1. □

We will now show that if conditional directed information, $I(W_i \to W_j \parallel \mathcal{W}_{\overline{ji}})$, are computed using corrupted data-streams, and were applied for causal structure inference, then we infer the perturbed graph that contains spurious links.

Corollary 2. *Consider a generative graph, $G = (V, A)$, consisting of N nodes. Let $Z = \{v_1, \ldots, v_n\} \subset V$ be the set of n perturbed nodes where each perturbation is described by (4.4). Denote the data-streams as follows: $U_Z := \{U_i\}_{i \in Z}$ and $Y_{\bar{Z}} := \{Y_j\}_{j \in \bar{Z}}$ where $\bar{Z} = V \setminus Z$. Let the measured data-streams be $\mathcal{W} = U_Z \cup Y_{\bar{Z}} = \{W_1, W_2, \ldots, W_N\}$. Let the perturbed graph be $G_Z = (V, A_Z)$. If $I(W_i \to W_j \parallel \mathcal{W}_{\overline{ji}}) > 0$, then $i \to j \in A_Z$.*

Proof. We will show that if $i \to j \notin A_Z$, then $I(W_i \to W_j \parallel \mathcal{W}_{\overline{ji}}) = 0$. Suppose, $i \to j \notin A_Z$. Let $G'_Z = (V', A'_Z)$ be the perturbed dynamic Bayesian network (DBN) associated with the perturbed graph, G_Z. Then, using Theorem 6, for all $t > 0$, d-sep$(W_j[t], W_i^{(t-1)} \mid W_i^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)})$ holds in G'_Z. In other words, this implies $P_{W_j[t]|W_j^{(t-1)}, W_i^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}} = P_{W_j[t]|W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}}$ will hold true for all t and thus, $I(W_i \to W_j \parallel \mathcal{W}_{\overline{ji}}) = 0$. □

The following example illustrates the intuition for the presence of spurious links in the perturbed graph.

Example 1. *Consider a generative graph as shown in Figure 4.4 (a). Suppose node* 3 *is subject to packet drop corruption model in* (4.7) *and let* U_3 *be its measured data-stream. Denote the measured data-streams at nodes 1 and 2 as* Y_1 *and* Y_2*.* U_3 *is related to its ideal counterpart* Y_3 *via* (4.8)*. The measured data-streams are* $\{W_1 = Y_1, W_2 = Y_2, W_3 = U_3\}$*. Since measurements of node 3 are corrupted, measurements of* Y_1 *and* Y_2 *can give useful information for predicting states at node 3 that would not be available in the noisily measured history of* U_3*. Thus,* $I(W_1 \to W_3 \parallel W_2) > 0$ *and* $I(W_2 \to W_3 \parallel W_1) > 0$*. The perturbed graph is shown in figure 4.4 (b).*

(a) True Graph, G. (b) Perturbed Graph, G_Z.

Figure 4.4: This figure illustrates the intuition behind spurious links in Example 1. Figure 4.4(b) shows the perturbed graph inferred. Spurious links are shown in red and the true edges are depicted in black.

The results in Theorem 6 and Corollary 2 respectively shows that existence of active trails is the PDBN and non-zero conditional directed information is sufficient to infer the presence of a directed link in the perturbed graph. However, under a mild assumption on the generative and the perturbation model, it can be shown that the respective conditions are also necessary to detect a directed link in the perturbed graph.

Assumption 1. *Let the following conditions on the generative and the perturbation model hold:*

C1) In the generative model (4.1)*, for all agents* $i \in \{1, 2, \ldots, N\}$*, and all* $j \in \mathcal{P}'(i)$*, there is a number* $k_{ij} \geq 1$ *such that* $Y_j[t - k_{ij}]$ *is an argument of* f_i*.*

C2) For all perturbed nodes $i \in Z$*, in the perturbation model* (4.4)*, there is a number* $k_i \geq 1$ *such that* g_i *always takes* $Y_i[t - k_i]$ *as it's argument.*

In addition, let at least one of the following conditions on corruption model hold:

B1) If a node $i \in Z$, then there is a number $k_i' \geq 1$ such that $Y_i[t - k_i']$ is an argument of f_i in (4.1).

B2) If a node $i \in Z$, then $Y_i[t]$ is an argument of g_i in (4.4).

Remark 8. *The above assumption states that the dynamics in generative model* (4.1)*, $Y_i[t]$ depends on at least one previous measurement value of its parent nodes. Similarly, for the perturbation model* (4.4)*, the corrupt value $U_i[t]$ depends causally on uncorrupted measurement value. We consider strictly causal interactions in the generative model and causal interactions in the corruption model and are therefore realistic in many practical physical systems.*

The following theorem asserts that if $i \to j \in A_Z$ then there exists a corresponding active trail in perturbed DBN. The proof is given in appendix C.2.

Theorem 7. *Consider a generative graph, $G = (V, A)$, consisting of N nodes. Let $Z = \{v_1, \ldots, v_n\} \subset V$ be the set of n perturbed nodes where each perturbation is described by* (4.4)*. Denote the data-streams as follows: $U_Z := \{U_i\}_{i \in Z}$ and $Y_{\bar{Z}} := \{Y_j\}_{j \in \bar{Z}}$ where $\bar{Z} = V \setminus Z$. Let the measured data-streams be $\mathcal{W} = U_Z \cup Y_{\bar{Z}} = \{W_1, W_2, \ldots, W_N\}$. Suppose, the generative model and the perturbation model satisfies the conditions for dynamics that is mentioned in Assumption 1. If there is a directed edge from i to j in perturbed graph, $G_Z = (V, A_Z)$, then there exists a trail between a node in $W_i^{(t-1)}$ and $W_j[t]$ that is active given $\{W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\}$ in G_Z', for some $t > 0$.*

Under the following assumption we can in fact show that $I(W_i \to W_j \parallel W_{\overline{ji}}) > 0$ is also a necessary condition for $i \to j \in A_Z$ as shown in Corollary 3.

Assumption 2. *We assume that the generative model in* (4.1) *and the perturbation model in* (4.4) *are such that the corresponding DBN and PDBN are faithful Bayesian networks. Moreover, we consider positive joint distributions for the random processes Y and U.*

Corollary 3. *Under Assumption 2 and dynamics following Assumption 1, if $i \to j \in A_Z$, then $I(W_i \to W_j \parallel W_{\overline{ji}}) > 0$.*

Proof. By theorem 7, if $i \to j \in A_Z$, then there exists an trail in PDBN between $W_i^{(t-1)}$ and $W_j[t]$ that is active given $\{W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\}$ in G'_Z, for some $t > 0$. Under faithfulness assumption, this implies $P_{W_j[t]|W_j^{(t-1)},W_i^{(t-1)},\mathcal{W}_{\overline{ji}}^{(t-1)}} \neq P_{W_j[t]|W_j^{(t-1)},\mathcal{W}_{\overline{ji}}^{(t-1)}}$. Thus, $I(W_i \to W_j \parallel \mathcal{W}_{\overline{ji}}) > 0$. □

Remark 9. *The faithfulness assumption is justified as the unfaithful probability distributions are restricted to a set of Lebesgue measure zero [36]. Here, system parameters for which the algebraic conditions for the conditional independence hold true with true dynamical dependencies must belong to the set of measure zero.*

4.5 Summary

We studied the problem of inferring directed graphs for a large class of networks that admit nonlinear and strictly causal interactions between several agents. Our analysis further reinforces that spurious links arising from data corruption is a more general phenomenon. We provided necessary and sufficient conditions that delineates the effects of data corruption on the directed network structure inferred using directed information. We presented a tight characterization for the spurious links that arise due to corruption of data-streams by determining their location and orientation. Our results serve as a necessary first step in understanding what part of network reconstruction can be trusted to facilitate accurate system identification.

Chapter 5

Estimation of Conditional Directed Information

5.1 Introduction

In the previous chapter, we provided necessary and sufficient conditions that delineates the effects of data corruption on the directed network structure inferred using directed information. In this chapter, a method for estimating measures of statistical dependency in networks of nonlinear dynamical interactions from the observed time-series data is developed. We develop a conditional directed information estimator for time-series admitting finite alphabets and establish consistency results for the estimator. Using this estimator, we verify the theoretical predictions of inferring and localizing spurious links in the directed graph using corrupt data streams.

Given the time-series $\mathcal{W}$, the reconstruction of the perturbed graph is accomplished by (i) computing the conditional directed information, $I(W_i \to W_j \parallel W_{\overline{ji}})$ (ii) placing a link from node i to j if $I(W_i \to W_j \parallel W_{\overline{ji}}) > 0$. Thus, the algorithm requires computation of $I(W_i \to W_j \parallel W_{\overline{ji}})$ for all pairs of nodes (W_i, W_j) in $\mathcal{W}$. Toward computing the conditional directed information we refer to methods based on Context-Tree-Weighting (CTW) in [181], which provide estimates on conditional probability mass function (PMF) of a time-series admitting values in a finite alphabet. For the reminder of this section, we consider f_i in (4.1) and g_i in (4.4) to be such that $Y_i[t]$ and $U_i[t]$ belong to finite alphabet. Here, from time sequence $x^{(n)}$ (recall the notation of

$x^{(n)}$), the PMF $Q(x[i] \mid x^{(i-1)})$ for all $i = 1, \ldots, n$ is computed where n is the length of the sequence. Q is also called as sequential probability assignment for a sequence $x^{(n)}$. Furthermore it is shown in [182] that Q computed is a *Universal Probability Assignment* as discussed next.

5.2 Universal Probability Assignment

The following definition characterizes the probability mass function Q in relation to the true mass function P in terms of the length of the time-series. It establishes that as the horizon of the time-series is extended, the sequential probability assignment estimate, Q, approaches the true PMF P.

Definition 28 (Universal Probability Assignment)**.** *Let P be the true joint PMF of $x^{(n)}$. Then, a probability assignment Q is called as* universal *if the following holds:*

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left[\log \frac{P(x^{(n)})}{Q(x^{(n)})}\right] = 0, \tag{5.1}$$

where estimated joint PMF for $x^{(n)}$ is given by $Q(x^{(n)}) = Q(x[0])Q(x[1] \mid x[0])Q(x[2] \mid x^{(1)}]) \cdots Q(x[n] \mid x^{(n-1)})$. Similarly, $P(x^{(n)})$ can be factorized.

For the rest of the article Q is estimated by CTW algorithm which is a universal probability assignment as discussed in [182] for each time series. The only assumptions made are that the sequences belong to a finite alphabet and are stationary and ergodic Markov sequences of a bounded order D. That is, for a Markov sequence X, $P(x[t]|x^{(t-1)}) = P(x[t]|x_{t-l}^{t-1})$ where $l \leq D$. The CTW algorithm uses a weighted distribution to take into account of all possible D-bounded Markov sources and estimates the sequential probability, $Q(x[t]|x^{(t-1)})$ for every symbol $x[t]$ given the past observations. The computational complexity of CTW algorithm is linear in horizon length n, of the sequences considered.

5.3 Pairwise Estimation of Directed Information

Here, a pairwise estimator of directed information between a pair of random process proposed by [182] is described. Let X and Y be jointly stationary and ergodic processes.

The directed information from X to Y can be expressed in terms of the entropy as follows:

$$I(X \to Y) = H(Y) - H(Y \parallel X) \tag{5.2}$$

where $H(Y) = \mathbb{E}[-\log P(Y)]$ and $H(Y \parallel X) = \mathbb{E}[-\log P(Y \parallel X)]$ denotes the entropy of Y and the causally conditioned entropy [183] respectively.
The directed information rate (DIR) from X to Y is defined as:

$$I_r(X \to Y) = \lim_{n\to\infty} \frac{1}{n} I(x^{(n)} \to y^{(n)}). \tag{5.3}$$

The directed information rate in (5.3) characterizes the directed information from X to Y in the limiting sense of the horizon being infinite. Let $H_r(Y) := \lim_{n\to\infty} \frac{1}{n} H(y^{(n)})$ and let $H_r(Y \parallel X) := \lim_{n\to\infty} \frac{1}{n} H(y^{(n)} \parallel x^{(n)})$. Thus, if $H_r(Y)$ and $H_r(Y \parallel X)$ converge, then I_r is convergent. That is,

$$I_r = H_r(Y) - H_r(Y \parallel X). \tag{5.4}$$

In [182], the following DIR estimator is defined:

$$\begin{aligned}\hat{I}(x^{(n)} \to y^{(n)}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid x^{(i-1)}, y^{(i-1)}) \cdot \log \frac{1}{Q(y[i] \mid y^{(i-1)})} \\ - \frac{1}{n} \sum_{i=1}^{n} \sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid x^{(i-1)}, y^{(i-1)}) \cdot \log \frac{1}{Q(y[i] \mid x^{(i-1)}, y^{(i-1)})}\end{aligned} \tag{5.5}$$

In [182], consistency results for estimating directed information (DI) between a pair of random processes from data was proposed. In this article we provide consistency results of the conditional directed information estimator by showing convergence in almost sure sense (denoted as P-a.s).

5.4 Estimation of Conditional Directed Information

Let X, Y, Z be jointly stationary and ergodic processes. The conditional directed information from X to Y conditioned on Z can be expressed in terms of the entropy as follows:

$$I(X \to Y \parallel Z) = H(Y \parallel Z) - H(Y \parallel X, Z). \tag{5.6}$$

The causally conditioned directed information rate (DIR) from X to Y now is defined as:

$$I_r(X \to Y \parallel Z) = \lim_{n\to\infty} \frac{1}{n} I(x^{(n)} \to y^{(n)} \parallel z^{(n)}). \tag{5.7}$$

Let $H_r(Y \parallel X, Z) := \lim_{n\to\infty} \frac{1}{n} H(y^{(n)} \parallel x^{(n)},\ z^{(n)})$. Thus, if $H_r(Y \parallel Z)$ and $H_r(Y \parallel X, Z)$ converge, then I_r is convergent. That is,

$$I_r = H_r(Y \parallel Z) - H_r(Y \parallel X, Z). \tag{5.8}$$

The conditional directed information estimator $\hat{I}(x^{(n)} \to y^{(n)} \parallel z^{(n)})$ is defined as under:

$$\hat{I}(x^{(n)} \to y^{(n)} \parallel z^{(n)}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)}) \cdot \log \frac{1}{Q(y[i] \mid y^{(i-1)}, z^{(i-1)})}$$
$$- \frac{1}{n} \sum_{i=1}^{n} \sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)}) \cdot \log \frac{1}{Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)})} \tag{5.9}$$

The following theorem establishes the consistency result in estimating conditional DIR as defined in (5.9). The proof is given in appendix D.1.

Theorem 8. *Let Q be the probability assignment in the CTW algorithm. Suppose, X, Y, Z are jointly stationary irreducible aperiodic finite-alphabet Markov processes whose order is bounded by the prescribed tree depth of the CTW algorithm. Then,*

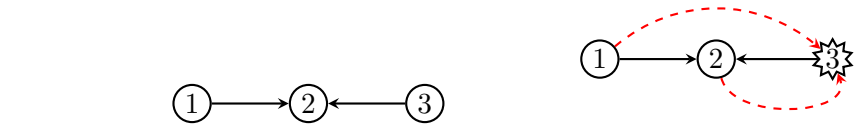
$$\lim_{n\to\infty} \hat{I}(x^{(n)} \to y^{(n)} \parallel Z^{(n)}) = I_r(X \to Y \parallel Z) \qquad P\text{-}a.s, \tag{5.10}$$

For computing (5.9), first $Q(x[i], y[i], z[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)})$ and $Q(y[i], z[i] \mid y^{(i-1)}, z^{(i-1)})$ are estimated using CTW for all realizations of tuples $(x[i], y[i], z[i])$ and $(y[i], z[i])$. The estimated probabilities are tabulated and the required marginalized conditional probabilities $Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)})$ and $Q(y[i] \mid y^{(i-1)}, z^{(i-1)})$ in (5.9) are computed from this table for entropy estimation.

5.5 Simulation Results

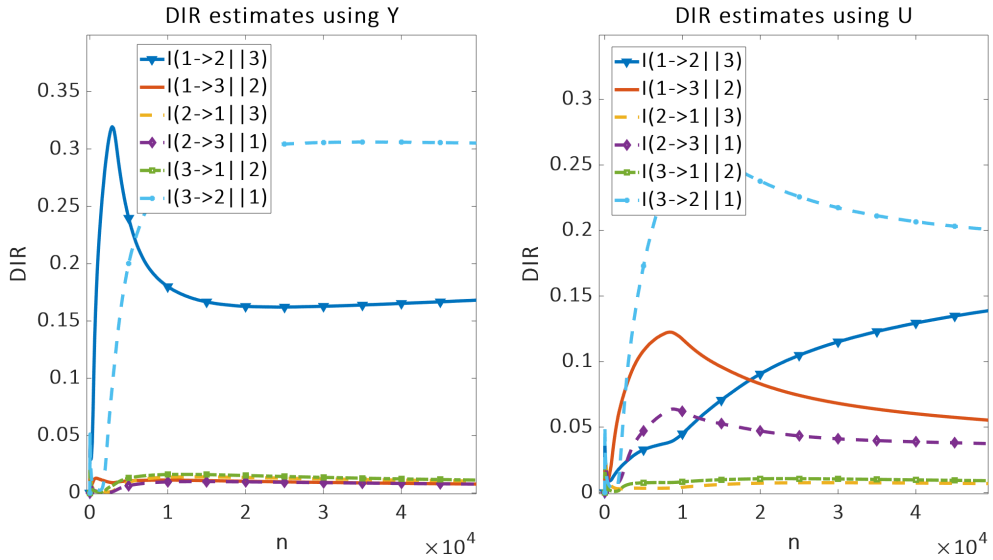
To verify the predictions of Theorem 6, we first performed a simulation on a network consisting of 3 nodes with a single node being perturbed and on a network consisting of 6 nodes, of which 2 are corrupt. We estimate the directed information rates (DIR),

which are DI estimates that are averaged along the sequence length until the horizon. We used the estimator described in (5.9) to compute DIR. For both the networks, the horizon length are in the order 10^4.



(a) True Graph

(b) Perturbed Graph.



(c) DIR Estimates comparison

Figure 5.1: This figure shows how unreliable measurements at node 3 results in spuriously inferring a causal influence from $1 \to 3$ and $2 \to 3$. 5.1(b) shows the perturbed graph inferred. Spurious edges are shown in red while true edges are in black.

5.5.1 Single node Perturbation

Consider a network consisting of 2 nodes with a common child as shown in Fig. 5.1(a). The true generative model is described as follows:

$$\begin{aligned} Y_1[t] &= E_1[t], \\ Y_2[t] &= (|Y_1[t-1] - Y_1[t-2]| \cdot Y_3[t-1]^2 \cdot E_2[t]) \bmod 3, \\ Y_3[t] &= E_3[t] \end{aligned}$$

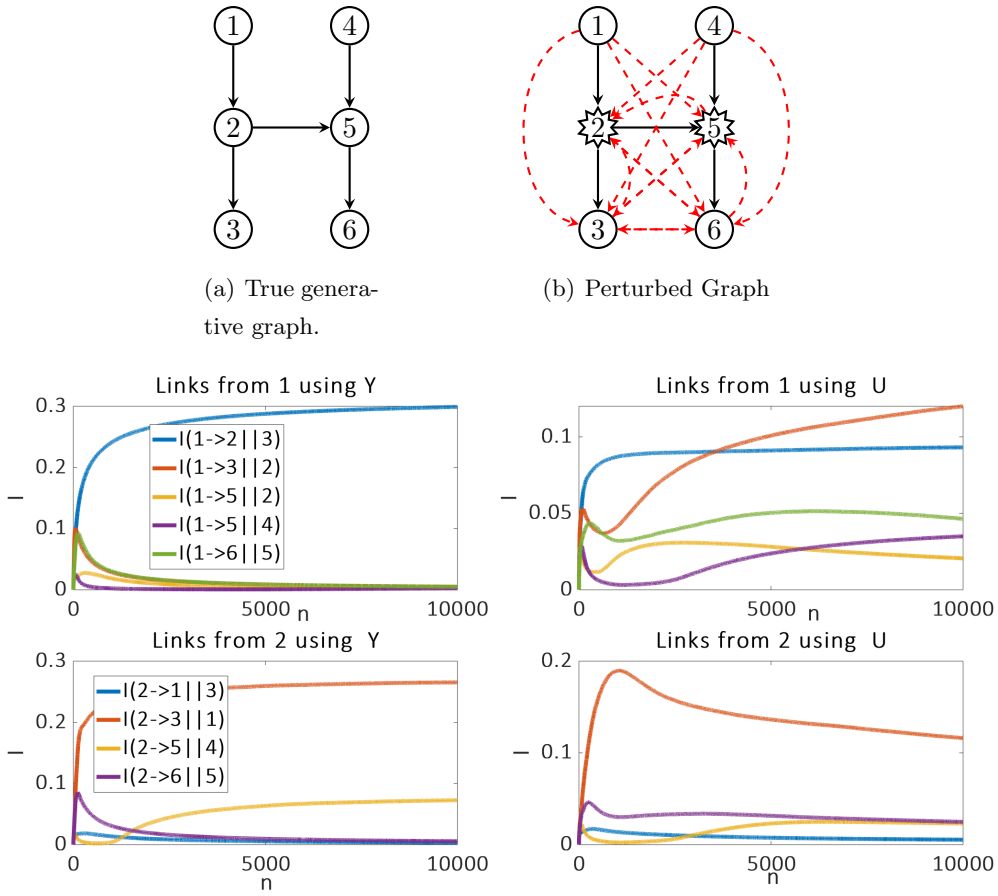
where $E_1[t] \sim \text{Categorical}(3, [0.15,\ 0.35,\ 0.5])$, $E_2[t] \sim \text{Categorical}(3, [0.35,\ 0.35,\ 0.3])$ and $E_3[t] \sim \text{Categorical}(3, [0.4,\ 0.2,\ 0.4])$. Each of $Y_1[t], Y_2[t]$ and $Y_3[t]$ has a finite alphabet $\{0, 1, 2\}$.

The perturbation considered here is the packet-drops uncertainty at node 3. The corruption model takes the form:

$$U_3[t] = \begin{cases} (Y_3[t] + U_3[t-1]) \bmod 3, \text{ with probability } 0.55 \\ (Y_3[t-1] + U_3[t-1]) \bmod 3, \text{ with probability } 0.45. \end{cases}$$

The perturbed graph predicted by Theorem 6 is shown in Fig. 5.1(b). The DIR estimates from ideal (Y) and unreliable measurements (U) are shown in Fig. 5.1(c). We observe non-zero DIR estimates and add edges to G_Z respectively. In particular, note the substantial rise in $I(U_1 \to U_3 \parallel U_2)$ and in $I(U_2 \to U_3 \parallel U_1)$. This indicates the presence of spurious links $1 \to 3$ and $2 \to 3$ in the inferred perturbed graph.

5.5.2 Multiple Perturbation



(a) True generative graph.

(b) Perturbed Graph

(c) DIR Estimates

Figure 5.2: 5.2(a) shows true generative graph. 5.2(c) depicts DIR estimates to detect links from nodes 1 and 2 using ideal measurements Y and when there is corruption at nodes 2 and 5. 5.2(b) shows the perturbed graph inferred. The spurious links are shown in red and the true edges are shown in black. With cascaded perturbations, more spurious links are inferred.

Consider a network of 6 nodes as shown in Fig. 5.2(a). The dynamic interactions in the true generative model are as follows:

$$\begin{aligned} Y_1[t] &= E_1[t], \\ Y_2[t] &= Y_1[t-1] || E_2[t], \\ Y_3[t] &= Y_2[t-1] || E_3[t], \\ Y_4[t] &= E_4[t], \\ Y_5[t] &= (Y_2[t-1] || Y_4[t-1]) \& E_5[t], \\ Y_6[t] &= Y_5[t-1] || E_6[t] \end{aligned}$$

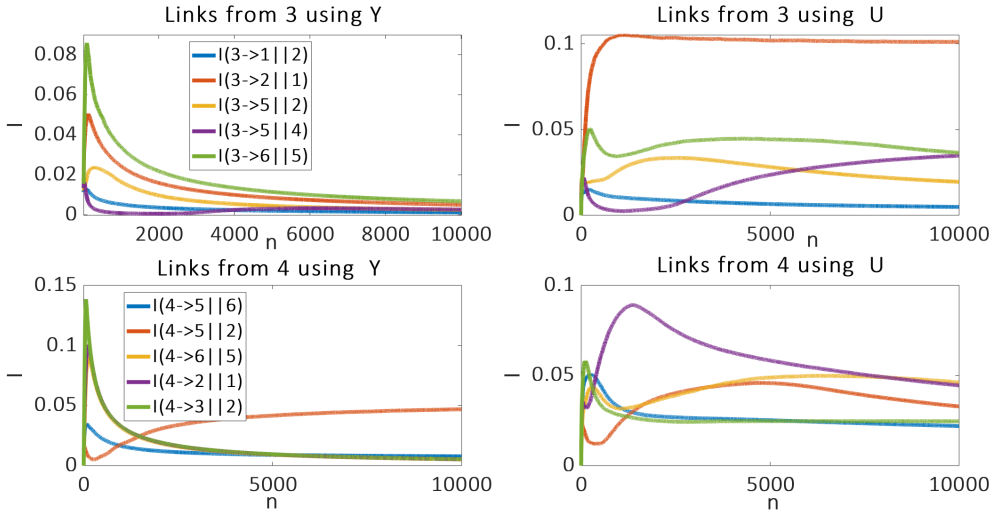
where $E_1[t] \sim$ Bernouilli(0.55), $E_2[t] \sim$ Bernouilli(0.5), $E_3[t] \sim$ Bernouilli(0.2), $E_4[t] \sim$ Bernouilli(0.4), $E_5[t] \sim$ and $E_6[t] \sim$ Bernouilli(0.3)and '||' is logical 'OR' operation while '&' is logical 'AND' operation. Each of $Y_1[t], Y_2[t], \ldots, Y_6[t]$ has a finite alphabet $\{0, 1\}$. The perturbations considered here are time-origin uncertainties at nodes 2 and 5. The corruption models takes the form:

$$U_2[t] = \begin{cases} Y_2[t-2], & \text{with probability 0.5} \\ Y_2[t], & \text{with probability 0.5.} \end{cases}$$

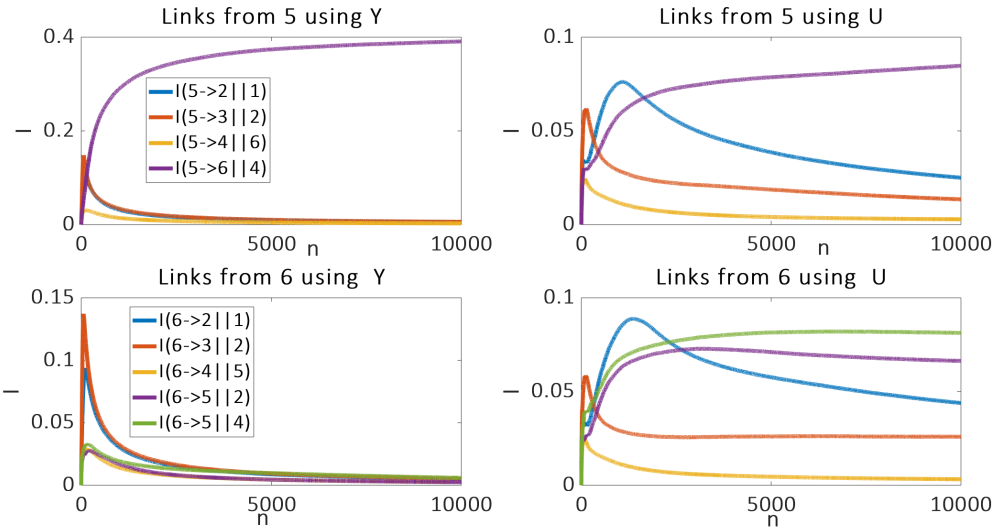
and

$$U_5[t] = \begin{cases} Y_5[t-2], & \text{with probability 0.5} \\ Y_5[t], & \text{with probability 0.5.} \end{cases}$$

The perturbed graph predicted by Theorem 6 is shown in figure 5.2(b). The DIR estimates from ideal (Y) and unreliable measurements (U) are shown in figures 5.2(c) and 5.3. We observe non-zero DIR estimates and add edges to G_Z respectively. For clarity of visualization, only non-zero DIR estimates predicted by Theorem 6 are shown.



(a) DIR estimates



(b) DIR estimates from f and 6

Figure 5.3: DI estimates to detect links from nodes 3,4,5 and 6. Notice the large number of non-zero DIR estimates computed from corrupt measurements corresponding to links from nodes 3 and 6 which had no children in the true generative graph that now has lot of children nodes in G_Z.

5.6 Summary

We developed a method for estimating measures of statistical dependency in networks of nonlinear dynamical interactions from the observed time-series data. Specifically, a conditional directed information estimator for time-series admitting finite alphabets and finite Markov order is developed and convergence results for the estimator were established. Simulation results were provided to verify the theoretical predictions to infer and localize spurious links in the directed graph of nonlinear dynamical systems, using corrupt time-series.

Chapter 6

Corruption Detection in Networks of Bi-directional Dynamical Systems

6.1 Introduction

In this chapter, we show that our precise characterization of the spurious links from Chapter 4 can be further exploited when the network has special structural properties. There are several physical systems, especially flow-driven systems like power grid [116, 184], fluid flow [118,121] and heat transfer networks [122], where every dynamic coupling between the agents/nodes is bi-directional. Such a framework is also applicable in many engineered systems such as networks of oscillators [185] and consensus networks [186]. In such systems, identifying unidirectional links in reconstruction lead to the conclusion that such links arise from data corruption. Using our precise characterization of spurious links, we develop a procedure to detect the location of all corrupt nodes in bi-directed networks.

We consider causal and non-linear dynamical interactions between the agents. Directed information method is first employed to infer the *corrupt* graph from uncertain data-streams. We then use graph theory tools to isolate the corrupt nodes by observing the directed graph inferred. We remark here, that the solution methodology provides

an effective method to detect sources of corruption that only involve examining paths in the constructed graph.

6.2 Perturbed Graphs for Bi-directional Networks

In this section, we will provide a precise characterization of network structure inferred using corrupt data-streams for bi-directional networks. Recall the description of generative models as described in Chapter 4. In this chapter, we consider *bi-directional generative models* whose associated generative graph $G = (V, A)$, is bi-directional. That is for all $i \to j \in A$, we also have $j \to i \in A$. We use $i \rightleftharpoons j$ to denote $i \to j$ and $i \leftarrow j$. For an illustration, consider the dynamics of a generative model described by:

$$
\begin{aligned}
y_1[t] &= f_1(y_1^{(t-1)}, y_2^{(t-1)}, y_3^{(t-1)}, e_1[t]),\\
y_2[t] &= f_2(y_1^{(t-1)}, y_2^{(t-1)}, y_3^{(t-1)}, e_2[t]),\\
y_3[t] &= f_3(y_1^{(t-1)}, y_2^{(t-1)}, y_3^{(t-1)}, e_3[t]),
\end{aligned}
$$

Its associated generative graph is shown in Fig. 1(a). Note that for all i in $\{1, 2, 3\}$, $i \to i$ is not shown.

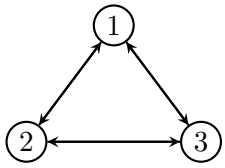


Figure 6.1: This figure shows a bi-directional generative graph.

We consider data corruption models and directed information based inference procedure described in Chapter 4. Recall that there are spurious links introduced in the inferred *perturbed graph*. The following proposition provides a precise and simplified characterization for perturbed graphs for networks whose generative graphs are bi-directional.

Proposition 1. *Let $G = (V, A)$ be a bi-directional generative graph. Let $Z \subset V$ be set of perturbed nodes and let $G_Z = (V, A_Z)$ be the corresponding perturbed graph. Then, $i \to j$ in A_Z if and only if either one of the following condition holds:*

B1) $i \rightleftharpoons j$ in G or

B2) There is a trail of length at least 3, $i = v_1 \rightleftharpoons v_2 \rightleftharpoons v_3 \rightleftharpoons \cdots \rightleftharpoons v_k = j$, such that for every pair of consecutive nodes v_m, v_{m+1} with $m \geq 2$ at least one of v_m or v_{m+1} is in Z.

6.3 Identification of Corrupt Nodes

We consider an important class of bi-directional networks whose operational topology or configuration is a tree. Such networks are prevalent across many applications such as electricity distribution grids [21], gas transmission networks [187], water networks [188], and information broadcasting [100]. In the previous section we gave a precise characterization of perturbed graphs for networks whose generative graphs are bi-directional. It is conceivable that if there are unidirectional links in the network structure inferred using data-streams, then there are corrupt nodes. We will now see how the location of unidirectional spurious links can be utilized to detect corrupt nodes. For the rest of the section we have the following assumption on the perturbations.

Assumption 3. *Let $G = (V, A)$ be a bi-directional generative graph. Let $Z \subset V$ be the set of perturbed nodes satisfying* (4.4)*. We consider perturbations that satisfy the following: for every unperturbed node $i \in V$ there exists at least one more unperturbed node $j \in V$ such that $i \rightleftharpoons j$ holds in G.*

Remark 10. *The above assumption states that we consider perturbations such that every unperturbed node in the generative graph is connected to at least one other unperturbed node. However, any node(corrupt/unperturbed) can be can be connected to multiple perturbed nodes. Thus, the above assumption is not restrictive.*

Theorem 9 is the main result which detects the exact location of all the corrupt nodes in the network. To this, we will require the following definitions.

Definition 29 (Topology)**.** *Suppose $G = (V, A)$ is a directed graph. The* topology *of the graph G is an undirected graph $G^\tau = (V, A^\tau)$, where $A^\tau = \{i - j \mid i \to j \in A\} \cup \{i - j \mid i \leftarrow j \in A\}$.*

Definition 30 (Tree)**.** *An undirected graph $G = (V, A)$ is called a* tree *if there is a unique path connecting any two nodes in V.*

Definition 31 (Bi-directional Clique)**.** *Suppose, $G = (V, A)$ is a directed graph. A subset of nodes $S \subset V$ forms a* bi-directional clique *in G if $i \to j \in A$ and $j \to i \in A$ for all $i, j \in S$.*

Definition 32 (Bi-directional Neighbors)**.** *Suppose, $G = (V, A)$ is a directed graph. The* bi-directional neighbors *of a node $i \in V$, $bidNr(i)$, is given by: $bidNr(i) = \{j \mid i \rightleftharpoons j$ holds in $G\}$.*

Theorem 9. *Suppose $G = (V, A)$ is a bi-directional generative graph with the topology, $G^\tau = (V, A^\tau)$, being a tree. Let $Z \subset V$ be the set of perturbed nodes satisfying* (4.4) *and Assumption 3. Let $U = \{u_1, \ldots, u_N\}$ be the measured data streams. Let $G_Z = (V, A_Z)$ be the perturbed graph. Consider the following statements:*

T1) $i \to j \in A_Z$

T2) $j \to i \notin A_Z$.

T3) j and $bidNr(j)$ form a bi-directional clique in G_Z.

If statements T1) and T2) hold, then j is corrupted if and only if statement T3) holds.

The following algorithm describes the procedure to detect all perturbed nodes in the network.

Algorithm 1 Detection of Corrupt Nodes

Input: Perturbed Graph, $G_Z = (V, A_Z)$.

Output: Set of perturbed nodes, Z.

Init: $Z \leftarrow \{\}$.

for all $i \to j \in A_Z$ **do**

  if $j \to i \notin A_Z$ **then**

    if $\forall p, q \in \text{bidNr}(j),\ p \rightleftharpoons q$ holds in G_Z **then**

      $Z \leftarrow Z \cup \{j\}$

    end if

  end if

end for

6.3.1 Illustrative Example

In this subsection, we provide examples and discuss the significance of the detection procedure described above.

Example 2 (Non-Linear System)**.** *Consider a network consisting of 3 nodes as shown in Figure 6.2(a)). The true generative model is described by:*

$$y_1[t] = e_1[t].y_2[t-1], \tag{6.1a}$$

$$y_2[t] = y_1[t-1] + y_3[t-1] \cdot e_2[t], \tag{6.1b}$$

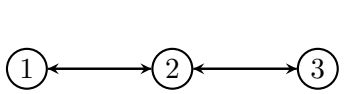
$$y_3[t] = y_2[t-1] + e_3[t] \tag{6.1c}$$

where $e_1[t] \sim$ Bernouilli(0.2), $e_2[t] \sim$ Bernouilli(0.2) and $e_3[t] \sim$ Bernouilli(0.35), and '$+$' is the logical 'OR' operation and '$\cdot$' is the logical 'AND' operation. Each of $y_1[t], y_2[t]$ and $y_3[t]$ has a finite alphabet $\{0, 1\}$. The perturbation considered here is the time-origin uncertainty at node 3. The corruption model takes the form:

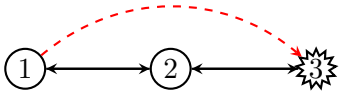
$$u_3[t] = \begin{cases} y_3[t-3], & \textit{with probability } 0.33 \\ y_3[t], & \textit{with probability } 0.67. \end{cases} \tag{6.2}$$

We used the methods proposed in [182] to compute directed information rate(DIR). The perturbed graph is shown in Figure 6.2(b)). Observe that only node 3 forms a

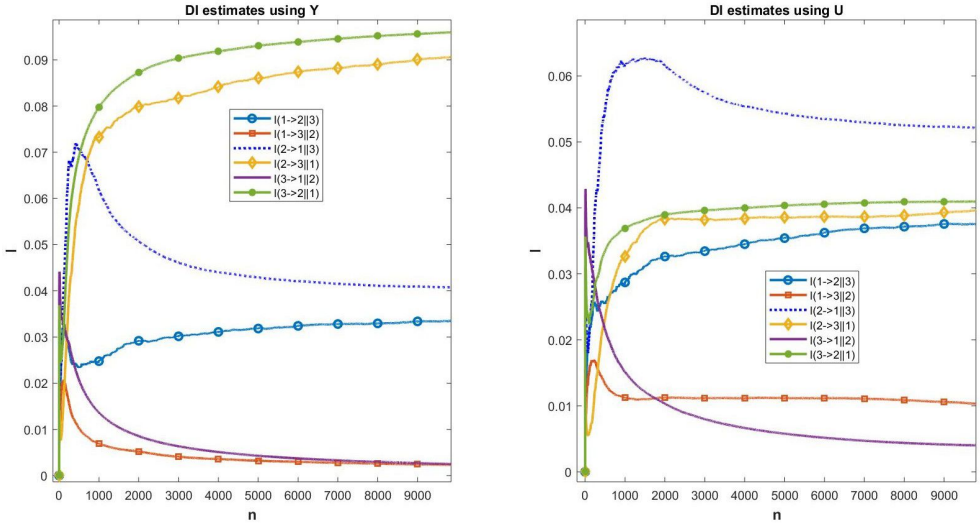
bi-directional clique with all its bi-directional neighbors. Thus, by Theorem 9, it is a perturbed node.



(a) Ideal Measurements Y

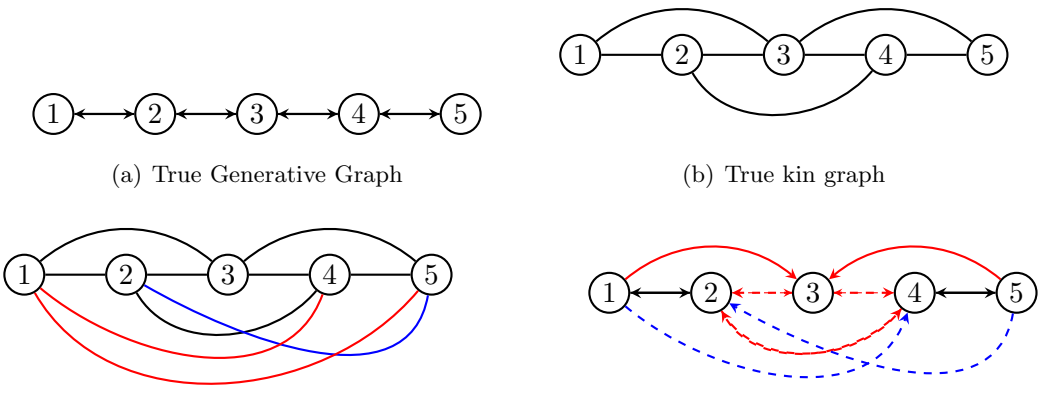


(b) Unreliable Measurements U.



(c) Comparison of DIR estimates between perfect measurements and corrupted data-streams.

Figure 6.2: **Directed Information estimates:** Note that only $I(y_3 \to y_1||y_2)$ and $I(y_1 \to y_3||y_2)$ keeps decreasing and asymptotically reaches zero. On the other hand, there is a rise in $I(u_1 \to u_3||u_2)$ and it plateaus as sample length grows while $I(u_3 \to u_1||u_2)$ continues to decrease. $I(u_2 \to u_3||u_1)$ shoots and plateaus at an increased value compared to $I(y_2 \to y_3||y_1)$ while other DI estimates continue to increase. We therefore infer the perturbed graph in Fig. 6.2(b)).



(a) True Generative Graph

(b) True kin graph

(c) Perturbed kin graph

(d) Directed perturbed graph

Figure 6.3: This figure shows how graph theory notions can detect corrupt nodes from the inferred directed perturbed graph. In 6.3(d)), node 3 forms bi-directional clique with all its bi-directional neighbors as shown in dashed red arrows. This does not hold for unperturbed nodes.

Example 3 (Linear System)**.** *Let the generative graph, G, be as shown in Fig. 6.3(a)) with the following dynamics:*

$$\begin{aligned}
y_1[t] &= (H_{12} * y_2)[t] + e_1[t] \\
y_2[t] &= (H_{21} * y_1)[t] + (H_{23} * y_3)[t] + e_2[t] \\
y_3[t] &= (H_{32} * y_2)[t] + (H_{34} * y_4)[t] + e_3[t] \\
y_4[t] &= (H_{43} * y_3)[t] + (H_{45} * y_5)[t] + e_4[t] \\
y_5[t] &= (H_{54} * y_4)[t] + e_5[t]
\end{aligned} \tag{6.3}$$

where H_{ij} are stable LTI filters for all $i, j \in \{1, 2, 3, 4, 5\}$. Suppose, node 3 is perturbed. That is, the measured data-stream for node 3, $u_3[t]$, is given by a noisy filter:

$$u_3[t] = (L * y_3)[t] + \zeta_3[t]$$

where, L is a stable first-order filter and $\zeta_3[t]$ is an IID Gaussian noise. Let $U[t] = \begin{bmatrix} y_1[t] & y_2[t] & u_3[t] & y_4[t] & y_5[t] \end{bmatrix}^\top$ be the vector of observed signals.

If the measurements were unperturbed, the true kin graph *(undirected graph) can be recovered using multivariate Wiener filter [74]. See Fig. 6.3(b)). However, if the*

measurements of y_3 are perturbed, then the undirected network structure identified using Wiener filtering will contain spurious links as proved in 2. In this case, the perturbed kin graph will be complete, as shown in Fig. 6.3(c). The Wiener filtering method implies that if $i - j$ is in the kin topology, then the corresponding entry in the inverse power spectrum satisfies $(\Phi_{UU}(e^{j\omega})^{-1})_{ij} \neq 0$ for all ω. From a trajectory of length 5×10^5, we estimated:

$$\Phi_{UU}(1)^{-1} = \begin{bmatrix} 1.052 & -1.233 & 0.007 & 0.242 & -0.073 \\ -1.233 & 1.621 & -0.036 & -0.538 & 0.227 \\ 0.007 & -0.036 & 0.028 & -0.034 & 0.005 \\ 0.242 & -0.538 & -0.034 & 1.515 & -1.146 \\ -0.073 & 0.227 & 0.005 & -1.146 & 0.983 \end{bmatrix}. \tag{6.4}$$

Thus, the Wiener filter method predicts a full graph. Due to high symmetry and completeness of the graph, it is not possible to detect the corrupt node purely by looking at the inferred graph structure. Moreover, separation techniques as described in [189] cannot be used to remove spurious edges.

However, directed perturbed graph yields more insightful results. This could be estimated from data using directed information [78], or in the case of linear systems, Granger causality [190].

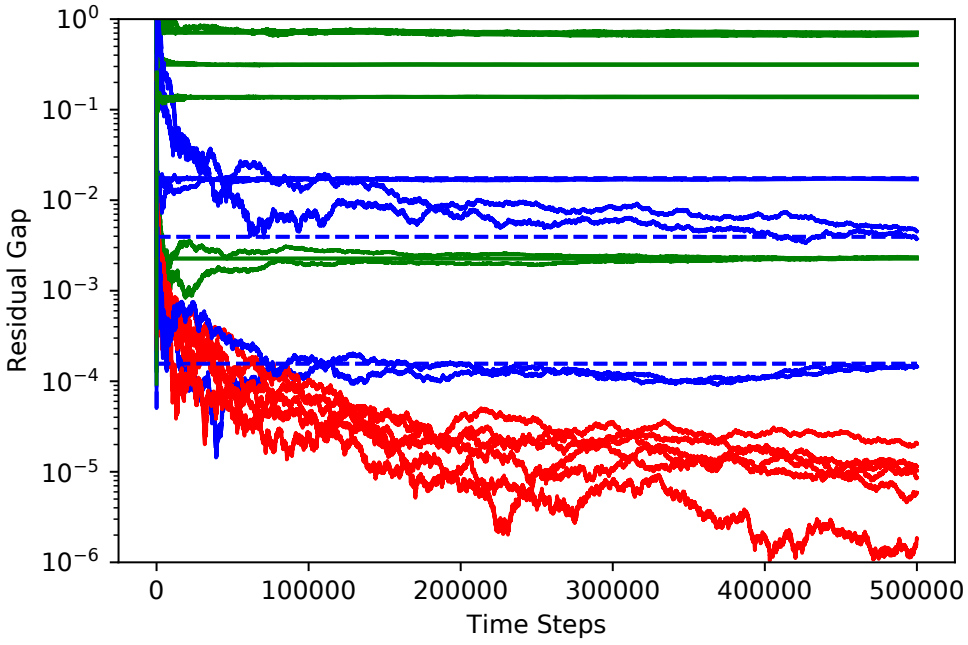


Figure 6.4: **Residual Errors**. Each line depicts a running estimate of the residual gap, (6.5). The dashed lines correspond to theoretical predictions, while the solid lines are the estimates. The green lines correspond to true links $i \to j \in A$, the blue lines correspond to spurious links $i \to j \in A_Z \setminus A$, and the red lines correspond to pairs (i, j) with no predicted link $i \to j \notin A_Z$. As can be seen the green and blue lines plateau near predicted values while the red lines continue to decrease.

As discussed in [190] Granger causality is equivalent to directed information in the case of linear Gaussian dynamic systems. Let $U_{\bar{j}}$ denote the entries of U other than j and let $U_{\overline{ij}}$ denote the entries of U other than i and j. For the Granger filter methods, we used recursive least-squares to estimate the difference in prediction error residuals:

$$r_{i,j} \quad = \quad \mathbb{E}\left[\left(u_j[t] - \mathbb{E}[u_j[t]|U_{\bar{j}}^{t-1}]\right)^2\right] \quad - \quad \mathbb{E}\left[\left(u_j[t] - \mathbb{E}[u_j[t]|U_{\overline{ij}}^{t-1}]\right)^2\right]. \quad (6.5)$$

It can be shown that for linear Gaussian systems that $r_{i,j} > 0$ if and only if $I(u_i \to u_j|U_{\overline{ij}}) > 0$. The results of the estimation are shown in Fig. 6.4. As can be seen, Fig. 6.3(d) is the inferred perturbed graph, G_Z. Applying Theorem 9, we can conclude node 3 as a perturbed node just by observing the graph structure.

6.4 Summary

We studied the problem of identifying the location of corrupt nodes in a network. We described a method to detect all the corrupt nodes in a network admitting causal and non-linear dynamical interactions with bi-directional coupling between its agents. We showed that solely by examining the perturbed graph structure we can locate corrupt nodes. In particular, we showed that a corrupt node will always form a bi-directional clique with all its bi-directional neighbors.

Chapter 7

Topology Learning in Radial Dynamical Systems with Unreliable Data

7.1 Introduction

Similar to the previous chapter, we consider dynamical systems where every coupling between agents/nodes is considered to be bi-directional. However, we focus on linear time-invariant dynamical relations between the agents. Moreover, unlike the problem setup in the previous chapter, we admit non-causal relations among agents and allow temporal correlations in the process noise. In the previous chapter, the dependencies in the generative model were restricted to be strictly causal and allowed only temporally uncorrelated process noise. Here, an undirected edge between two nodes may be effectively used to represent the bi-directional coupling between them. Thus, identifying the true Boolean structure (presence or absence of undirected edges) of the network correctly identifies all the influence pathways in the network. Moreover, we consider radial networks in this chapter. Radial networks constitute important class of networks that are prevalent across many applications such as electricity distribution grids [21], gas transmission networks [187], water networks [188], and information broadcasting [100]. Such systems follow tree network operational configuration.

7.1.1 Related Work

As discussed before, network identification for dynamical systems from time series measurements has been extensively studied in controls community [135], [74], [133], [108], [82], [105] [112]. In [189], the authors show that for bi-directed LTI system networks with a radial network topology, the spurious edges in the moral graph can be eliminated using graphical separation rules. Similarly, phase -based results of the estimated Wiener filters have been shown to enable removal of spurious edges in non-radial bi-directional LTI systems [86]. For a network of interacting agents with nonlinear dynamic dependencies, strictly causal interactions and temporally uncorrelated process noise, the authors in [78] proposed the use of directed information to determine the directed structure of the network. However, in this work we admit non-causal relations in the generative model and allow temporal correlations in the process noise. The problem of learning polytree structures has been studied in [80] and [85]. All the above work assume that the measurements are uniformly sampled and are available without any non ideal aspect like packet drops or random delays. Often, the data-streams in large systems are not immune to effects of noise [49], asynchronous sensor clocks [41], [45] and packet drops [54], [56]. In [179] focusing on directed networks with linear time-invariant (LTI) interactions, authors provided characterization of the extent of spurious links that can appear due to data-corruption in the moral graph. However, little is known if these spurious edges can be eliminated to infer the exact network structure even in the presence of corruptions in the data streams, thus establishing consistency guarantees. In [191], focusing on bi-directional networks, it is shown that the location of corrupt nodes can be detected by combining tools from information theory and graph theory. However, a method to eliminate spurious edges was not presented.

7.1.2 Our Contributions

In this chapter the objective is to determine the exact network representation of *radial* bi-directional LTI systems, using passive means from corrupt data-streams. Radial bi-directional LTI systems are characterized by a tree topology with bi-directed edges between neighbors. We show that for radial bi-directed network of LTI systems where corrupt nodes are located deep in the network, at least three hops away from the leaf

nodes, the spurious edges owning to data corruption can be eliminated and the exact network structure can be inferred. We present novel topological characterizations and phase-based properties to determine the exact location of corruptions. Finally, we propose an algorithm called 'hide and learn' to determine the exact topology generating the time series observations. To this, we follow a similar topology learning algorithm, presented in [127] that considered hidden nodes. However in [127], there was a tighter assumption on the distance between hidden nodes restricting them to be at least four hops away from each other and the measurements were assumed to be perfect. Here, we consider time-series with imperfect information and relax the assumption on the location of corrupt nodes, and provide rigorous proofs to our results.

The data corruption models considered are same as described in Section 2.3 of Chapter 2. In Section 7.3.1, we present the main result to determine the location of corrupt nodes. The exact topology learning algorithm is presented in Section 7.3.2.

7.2 Preliminaries

In this section the generative model and the generative graph that represents the networked system are presented.

7.2.1 Generative Model

Consider N agents that interact over a network. Consider the following continuous time linear dynamics for each agent $i \in \{1, \cdots N\}$:

$$\sum_{m=1}^{n} a_{m,i} \frac{d^m x_i}{dt^m} + a_{0,i} x_i(t) = \sum_{j=1, j\neq i}^{N} b_{ij} x_j(t) + w_i(t), \tag{7.1}$$

where the process $w_i(t)$ is considered to be zero mean WSS process innate to agent i and thus w_i is independent of w_j if $i \neq j$. Thus, the power spectral density (PSD) of $w = (w_1, w_2, \ldots, w_N)^T$, $\Phi_w(z)$, is a diagonal matrix. Above, $a_{m,i}$, $b_{ij} \in \mathbb{R}$ for all $m \in \{0, \ldots, n\}$ such that at least one $a_{m,i}$ is non-zero. We assume the signals are bounded in a mean-square sense: $\mathbb{E}[\| x_i[t] \|^2] < \infty$ and $\mathbb{E}[\| w_i[t] \|^2] < \infty$. Taking

Laplace transform of (7.1) we obtain:

$$\sum_{m=0}^{n} a_{m,i} s^m x_i(s) = \sum_{j=1, j\neq i}^{N} b_{ij} x_j(s) + w_i(s), \tag{7.2}$$

where $x_i(s)$ and $w_i(s)$ are the Laplace transform of $x_i(t)$ and $w_i(t)$ respectively. With Δt as sampling time and using bi-linear transformation of $s = \frac{2(1-z^{-1})}{\Delta t(1+z^{-1})}$, we obtain the following z transform of discretization of (7.1):

$$S_i(z) x_i(z) = \sum_{j=1, j\neq i}^{N} b_{ij} x_j(z) + w_i(z) \quad \text{for } i = 1, \ldots, N. \tag{7.3}$$

Here, $S_i(z) = \sum_{m=0}^{n} a_{m,i} \left(\frac{2(1-z^{-1})}{\Delta t(1+z^{-1})}\right)^m$ is the z domain operator determined by the derivatives of $x_i(t)$. Note that other discretization methods could also be used and our results are independent of the method. Rewriting (7.3) we obtain:

$$x_i(z) = \sum_{j=1, j\neq i}^{N} \mathcal{G}_{ij}(z) x_j(z) + e_i(z) \quad \text{for } i = 1, \ldots, N. \tag{7.4}$$

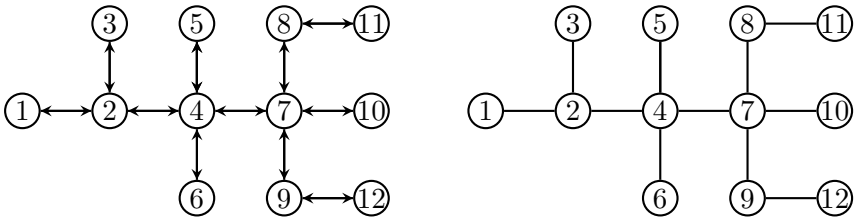
Here, $\mathcal{G}_{ij}(z) = \frac{b_{ij}}{S_i(z)}$, $e_i(z) = \frac{w_i(z)}{S_i(z)}$.

Compactly, (7.4) is equivalent to

$$x = \mathcal{G}(z)x + e, \tag{7.5}$$

where $x = (x_1(z), x_2(z), \ldots, x_N(z))^T$ and $e = (e_1(z), e_2(z), \ldots, e_N(z))^T$ and $\mathcal{G}(i,j) = \mathcal{G}_{ij}(z)$. We call the pair $(\mathcal{G}(z), e)$ the *generative model*. We consider bi-directed generative models such that $\mathcal{G}_{ij}(z) \neq 0$ if and only if $\mathcal{G}_{ji}(z) \neq 0$. Such models are prevalent in linearized models of engineering systems operating around an equilibrium point. For example, consider swing dynamics for power systems and heat transfer dynamical systems.

Remark 11. *Note that the accuracy of discretization will affect the sparsity of the network structure in the discrete time model. The sampling time, Δt, must commensurate with the time constants and dynamics of the system for the sampling-related non-zero entries to be negligible. For the rest of the chapter, we assume that the signals are sampled at high resolution.*



(a) Generative Graph G (b) Generative Topology G^T

Figure 7.1: A generative graph and its tree topology.

7.2.2 Graphical Representation

The structural description of (7.4) induces a *generative graph* $G = (V, \overrightarrow{A})$ formed by identifying the set of vertices, $V = \{1, 2, \ldots, n\}$, with random processes x_i and the set of directed links, $\overrightarrow{A}$ given by: $\overrightarrow{A} = \{i \to j | \mathcal{G}_{ji} \neq 0\}$. For bi-directional dynamical systems, it follows that $i \to j \in \overrightarrow{A}$ if and only if $j \to i \in \overrightarrow{A}$. . Thus, G is a bi-directed graph. Given the generative graph G for a bi-directional system, its *generative topology* is the undirected graph $G^T = (V, A)$ formed by replacing all directed edges in $\overrightarrow{A}$ with undirected edges such that multiple undirected edges between any two nodes is replaced by a single undirected edge between those same nodes. Here, the set $A = \{(i, j) \mid i \to j \in \overrightarrow{A}\} \cup \{(i, j) \mid j \to i \in \overrightarrow{A}\}$. The following definitions on undirected graphs will be useful for subsequent analysis. Figure 7.1 represents a bi-directed system.

Definition 33 (Path)**.** *Nodes $w_1, w_2, \ldots, w_n \in V$ forms a path in an undirected graph, $G = (V, A)$, if for every $i = 1, 2, \ldots, n-1$ we have $w_i - w_{i+1}$ in A. The path is denoted by $w_1 - w_2 \cdots - w_n$. The length of the path is one less than the number of nodes in the path.*

Definition 34 (n-Hop Neighbor)**.** *Given an undirected graph, $G = (V, A)$, a node $j \in V$ is a n-hop neighbor of $i \in V$ if there is a path of length n between i and j in G. We will denote n hop neighbors of i as $n - hop(i)$. We refer 1-hop neighbors as neighbors.*

Definition 35 (Leaf Node/Non-leaf Nodes)**.** *In a tree, $G^T = (V, A)$, a node $i \in V$ that has only one neighbor is called a leaf node. Nodes with more than one neighbor are called non-leaf nodes.*

Definition 36 (Radial Systems)**.** *If the generative topology, G^T, associated with a bi-directional generative model, $(\mathcal{G}(z), e)$, is a tree, then the generative system is called a* radial system*.*

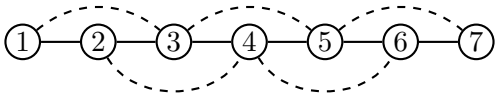
Figure 7.1 represents a radial system.

7.2.3 Moral/Kin Graph Inference from Time Series

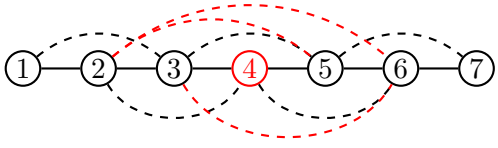
In Chapter 2, we discussed the relationship between the sparsity pattern of inverse PSD matrix of time sereis, x and the moral graph, G^M of a generative model. For a radial system, we can express Φ_{xx}^{-1} more explicitly, where the non-zero entries only correspond to diagonal entries, neighbors and 2-hop neighbors in G_T. For a radial system:

$$\Phi_{xx}^{-1}(i,j) = \begin{cases} -\mathcal{G}_{ij}(z)\Phi_{e_i}^{-1} - \mathcal{G}_{ji}(z^{-1})\Phi_{e_j}^{-1}, j \in 1-hop(i) \\ \mathcal{G}_{ki}(z^{-1})\mathcal{G}_{kj}(z)\Phi_{e_k}^{-1}, j \in 2-hop(i) \\ \qquad \text{and } k \in 1-hop(i), k \in 1-hop(j) \\ \Phi_{e_i}^{-1} + \sum_{k \in 1-hop(i)} |\mathcal{G}_{ki}|^2 \Phi_{e_k}^{-1}, i = j \\ 0, \text{otherwise.} \end{cases} \tag{7.6}$$

(a) Generative graph



(b) Kin graph inferred using perfect measurements



(c) Perturbed graph inferred with corruption at node 4

Figure 7.2: This figure shows how unreliable measurements at a node can yield in erroneous dynamic influences.

7.3 Exact Topology Learning from Corrupt Data Streams

The first step towards exact topology learning is to determine the location of all the corrupt nodes. We consider data perturbation models as discussed in Chapter 2. Consider a chain network consisting of 7 nodes with bidirectional dynamics between adjacent nodes as shown in 2.2(a). The true moral graph is depicted in figure 7.2(b). Suppose 4 is corrupted. Applying, Theorem 3, the inferred undirected graph is shown in figure 7.2(c). We consider the following assumption on the location of corrupt nodes:

Assumption 4. *C1) Corrupt nodes are at least 3 hops away from all leaf nodes in the generative topology.*

C2) Corrupt nodes are at least 3 hops away from each other in the generative topology.

Remark 12. *The above condition C1) implies that the corrupt nodes are located deep in the network such that its effects are felt by the agents that have perfect measurements. Moreover, C1) implicitly implies that every node has a 2-hop neighbor in the generative topology. We consider arbitrarily large networks satisfying this, and with at least one corrupt node. In particular, the smallest network permitted is a 7 node chain as shown in Figure 7.2(a)) with node* 4 *being corrupted.*

7.3.1 Corruption Detection

A method to locate the corrupt nodes in the inferred perturbed graph for radial dynamical systems is presented in this section. To this, we first characterize the topological properties of the neighborhood set of leaf and corrupt nodes in the inferred perturbed graph.

Neighborhood characterization

The following proposition states that that if i is a leaf or a corrupt node, its neighborhood set in the inferred perturbed graph, G_U, is completely determined by the set of 1-hop and 2-hop neighbors of i in the generative topology, G_T.

Proposition 2. *Consider a radial system with generative topology $G^T = (V, A)$ consisting of N nodes with the moral graph $G^M = (V, A^M)$. Let $1-hop(i)$ and $2-hop(i)$ denote*

the set of 1-hop and 2 hop neighbors of i in G^T. Let $Y \subset V$ be the set of perturbed nodes where each perturbation satisfies (2.10) *and Assumption 4. Suppose $G^U = (V, A^U)$ is the perturbed graph inferred using Theorem 3. Let neighbors of node i in G^U be $\mathcal{N}_u(i)$. If i is a leaf node in G^T or $i \in Y$, then $\mathcal{N}_u(i) = 1 - hop(i) \cup 2 - hop(i)$.*

Proof. We will show that no additional nodes excluding $1-hop(i)$ or $2-hop(i)$ neighbors exist as neighbors of i in G^U. Suppose j is a neighbor of i in G^U such that $j \notin 1 - hop(i) \cup 2 - hop(i)$. Then, by definition of perturbed graph, there should be a path $i - v_1 - v_2 - j$ in G^M such that v_1 and v_2 are corrupt nodes. This implies v_1 belongs to $1-hop(i)$ or $2-hop(i)$ in G^T. Suppose i is a corrupt node. This contradicts condition C2). Suppose i is a leaf node. This contradicts condition C1). Therefore, $\mathcal{N}_u(i) = 1 - hop(i) \cup 2 - hop(i)$. □

We will then use the following lemma that describes a topological method to detect a set of "candidate" nodes which contains only leaf and corrupt nodes using the perturbed graph. It states that only leaf nodes and corrupt nodes has a neighborhood that forms a clique in the perturbed graph. The proof is given in Appendix F.1.

Lemma 2. *Consider a radial dynamical system with generative topology $G^T = (V, A^T)$. Let $Y \subset V$ be the set of perturbed nodes where each perturbation satisfies* (2.10) *and Assumption 4. Suppose $G^U = (V, A^U)$ is the perturbed graph inferred using theorem 3. Consider any node i in V. Neighbors of node i in G^U, $\mathcal{N}_u(i) \cup \{i\}$ will form a clique in G^U if and only if i is a leaf node in generative topology G^T or i is a corrupt node.*

Detection of corrupt nodes

After a "candidate" set containing corrupt nodes and leaf nodes are determined as discussed above, we will now isolate the corrupt nodes exactly. The following theorem precisely detects the corrupt nodes separately based on phase properties of entries in inverse PSD. The proof is given in Appendix F.2

Theorem 10. *Suppose $G^T = (V, A^T)$ is the generative topology corresponding to a radial dynamical system. Let $Y = \{v_1, v_2, \ldots, v_n\} \subset V$ be set of corrupt nodes with each corruption satisfying* (2.10) *and Assumption 4. Suppose $G^U = (V, A^U)$ is the perturbed graph inferred using theorem 3. Let B be the set of nodes detected using Lemma 2 whose*

neighborhood, $\mathcal{N}_u(i)$ forms a clique with $\{i\}$ in G^U. Take a node $i \in B$. Then, i has at least two neighbors, p, q in G^U with non-constant $\angle\Phi_{uu}^{-1}(i,p)(\omega)$ for all $\omega \in (-\pi, \pi]$ if and only if i is a corrupt node in G^T.

The above result detects the set of corrupt nodes, y, from the candidate set, B, and hence the remaining nodes $B \setminus y$ are the leaf nodes. Further, Theorem 10 delineates that only leaf nodes have one unique entry in Φ_{uu}^{-1} with a non-constant phase response. This corresponds to the true edge associated with the leaf node. Let $\mathcal{E}_L$ be this set of edges. Thus, in addition to corrupt nodes, we also detect leaf nodes and remove spurious edges associated with leaf nodes applying Theorem 10. The procedure is described comprehensively in Algorithm 2.

Algorithm 2 Detection of Corrupt Nodes and Isolating True Edges Associated with Leaf Nodes

Input: Time series measurements, u.
Output: Set of perturbed nodes, Y, set of leaf nodes, L, and set of true edges, $\mathcal{E}_L$ associated with leaf nodes.
Init: $A_Z \leftarrow \{\}$, $Y \leftarrow \{\}$, $L \leftarrow \{\}$, $\mathcal{E}_L \leftarrow \{\}$.

Compute inverse PSD, Φ_{uu}^{-1}.
for all $i \in V$, $i \neq j$ **do**
  if $\Phi_{uu}^{-1}(i,j)(\omega) \neq 0$ **then**
    $A_Z \leftarrow A_Z \cup \{i-j\}$
  end if
end for
for all $i \in V$ **do**
  $\epsilon_i \leftarrow \{\}$.
  if $\mathcal{N}_u(i) \cup \{i\}$ forms a clique in $G_Z = (V, A_Z)$ **then**
    for all $j \in \mathcal{N}_u(i)$ **do**
      if $\angle\Phi_{uu}^{-1}(i,j)(\omega)$ is not constant for all $\omega \in (-\pi, \pi]$ **then**
        $\epsilon_i \leftarrow \epsilon_i \cup \{i-j\}$
      end if
    end for
    if Cardinality of $\epsilon_i \geq 2$ **then**
      $Y \leftarrow Y \cup \{i\}$
    else $L \leftarrow L \cup \{i\}$ and $\mathcal{E}_L \leftarrow \mathcal{E}_L \cup \epsilon_i$
    end if
  end if
end for

7.3.2 Hide and Learn Algorithm

The steps to recover the exact topology of the radial linear dynamical system using imperfect information are presented in this section. To accomplish this we follow *hide and learn* strategy. This is described in Algorithm 3 that has three stages: First, hide the measurements of the corrupt nodes. We infer the graphical structure of the network by observing sparsity pattern of inverse PSD using only the nodes that has perfect information by marginalizing out the corrupt node measurements. That is, the corrupt nodes will be treated as latent nodes. This graph will contain spurious edges. This constitutes lines 1 to 7 in Algorithm 3. Second, identify the true edges in the graph obtained from previous step. This constitutes lines 8 to 15 in Algorithm 3. Finally, place the corrupt nodes back at the correct location in the structure resulting from previous step as described in lines 16 to 29 in Algorithm 3. We now elaborate on each stage and provide technical results that support and yield Algorithm 3.

Learning with Latent Corrupt Nodes:

Let y be the time series measurements of corrupt nodes, Y, detected after Theorem 10. Let o denote the set of measurements without y. That is, $o = u \setminus y$. We compute the inverse PSD of o. Now, using sparsity pattern in inverse PSD of o as adjacency matrix construct an undirected graph, $\mathcal{T}_m = (V_o, A_o)$. The following result from [127] characterizes the edges in $\mathcal{T}_m$ and the generative topology G^T.

Lemma 3. *Consider a radial dynamical system with generative topology G^T. Then, $\Phi_{oo}^{-1}(i,j)(\omega) \neq 0$ for $\omega \in (-\pi, \pi]$, implies that i and j are within four hops of each other in G^T.*

True edge set discovery between observed nodes:

The graph $\mathcal{T}_m$ inferred from Lemma 3 contains spurious edges. The objective here is to eliminate the spurious edges and thus identify the true edges.
The following result from [127] provides a topological method based on separation property to identify the observed non-leaf nodes and identify the true edges between them.

Lemma 4. *Suppose $\mathcal{T}_m$ is the graph inferred using measurements o in Lemma 3. Suppose there exist observed nodes c, d distinct from observed nodes a, b such that $a - b \in$*

$\mathcal{T}_m$. Then, sep$(c, d|\{a, b\})$ holds in $\mathcal{T}_m$ if and only if $a - b$ is a true edge in G^T and a, b are non-leaf nodes.

Combining Lemma 4 with the output of Algorithm 2 that detected the only true edge associated with all the leaf nodes, we have thus identified all true edges associated with the observed nodes. Denote this graph as Θ.

Placement of Corrupt Nodes

The graph Θ will have multiple radial *disconnected components* denoted as θ_j, with the disconnections being at the location of the latent corrupt nodes, Y. Based on our assumptions, it can be shown that each disconnected component has at least two observed nodes. Thus, for all node $p \in \theta_j$, there is another node $q \in \theta_j$ such that $p - q \in G^T$. Since G^T is a connected graph, the final step is to connect the disconnected components by placing the corrupt nodes at the disconnected locations. We make use of the prior knowledge gained by inferring the perturbed graph G^U and we map every corrupt node $i \in Y$ to it's corresponding neighborhood $\mathcal{N}_u(i)$ in G^U. The following lemma precisely characterizes this. The proof is given in the Appendix F.3.

Lemma 5. *Let Θ be the disconnected network inferred after removing spurious edges between the observed nodes based on Lemma 4. Consider two disconnected components θ_1, θ_2 in Θ with observed nodes $q \in \theta_1$ and $r \in \theta_2$. Consider all $p \in \theta_1$ and all $s \in \theta_2$ such that $p - q$ and $r - s$ are edges in θ_1 and θ_2 respectively. Consider a corrupt node $l \in Y$. Suppose $\{p, q, l, r, s\}$ forms a clique in the perturbed graph, G^U. Then, $p - q - l - r - s$ holds in G^T if and only if $\angle\Phi_{uu}^{-1}(\omega)(p, s)$ is a constant (equal to 0 or π) for all $\omega \in (-\pi, \pi]$.*

Theorem 11 is the main result of the chapter which states that Algorithm 3 precisely learns the exact topology of a radial system with imperfect information once the corrupt nodes have been detected using Algorithm 2. The proof is given in the Appendix F.3.

Theorem 11. *Suppose Y is the set of perturbed nodes, L is the set of leaf nodes and $\mathcal{E}_L$ is the set of true edges associated with leaf nodes detected from Algorithm 2. Then, Algorithm 3 results in learning the true generative topology is $G^T = (V, A)$ for the corresponding radial system.*

Algorithm 3 Exact Topology Learning: *Hide and Learn*

Input: Inputs and outputs from Algorithm 2.

Output: Set of true edges, A, in generative topology G^T.

Init: Set of observed edges, $A_o \leftarrow \{\}$.

Isolate non-corrupt measurements, $o = u \setminus y$. Observed nodes $V_o = V \setminus Y$.

Using measurements o, compute inverse PSD, Φ_{oo}^{-1}.

for all $i \in V_o$, $i \neq j$ **do**

  if $\Phi_{oo}^{-1}(i,j)(\omega) \neq 0$ **then**

    $A_o \leftarrow A_o \cup \{i - j\}$

  end if

end for

Non-leaf nodes, $V_{nl} = V \setminus L$. True edge set, $\mathcal{E}_T \leftarrow \mathcal{E}_L$.

for all $p, q \in V_{nl}$ such that $(p,q) \in A_o$ **do**

  if There exist $K \neq \{\}$ and $S \neq \{\}$ such that $sep(K, S|\{p,q\})$ holds **then**

    $\mathcal{E}_T \leftarrow \mathcal{E}_T \cup \{(p,q)\}$, $A \leftarrow A \cup \mathcal{E}_T$

  end if

end for

$d \leftarrow$ number of disconnected components in the graph, $\Theta = (V_o, \mathcal{E}_T)$. (i.e $\Theta = \cup_{i=1}^{d} \theta_i$).

for all $i \in \{1, 2, \ldots, d\}$ and $j \in \{i+1, \ldots, d\}$ **do**

  if There exists nodes $q \in \theta_i$ and $r \in \theta_j$ such that $(p,q) \in \theta_i$ and $s - r \in \theta_j$ holds for some other observed nodes p, s **then**

    for all $l \in Y$ **do**

      if $\{p, q, l, r, s\}$ forms a clique in G_Z and if $\Phi_{uu}^{-1}(p,s)(\omega)$ is constant for all $\omega \in (-\pi, \pi]$ **then**

        $A \leftarrow A \cup \{q - l, l - r\}$

      end if

    end for

  end if

end for

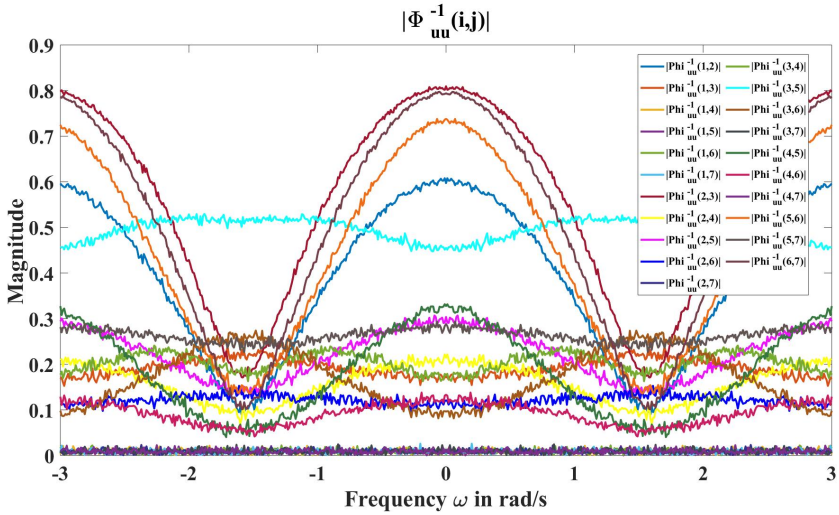


Figure 7.3: **Magnitude Plots** The magnitude of inverse power spectral density estimates computed from corrupt data streams u are shown in the here. Notice the entries are non-zero across the frequency grid. To each non-zero entry, we add undirected edges to infer the perturbed graph as shown in Figure 7.2(c)) following theorem 3.

7.4 Simulation Result

In this section we demonstrate the topological learning algorithm via a numerical example. Let the true generative graph, G, be as shown in Fig. 7.2(a)) with the following dynamics:

$$
\begin{aligned}
x_1[t] &= 0.5x_2[t-1] + e_1[t] \\
x_2[t] &= 0.36x_1[t-1] + 0.6x_3[t-1] + e_2[t] \\
x_3[t] &= 0.95x_2[t-1] - 1.7x_4[t-1] + e_3[t] \\
x_4[t] &= 0.51x_3[t-1] + 0.55x_5[t-1] + e_4[t] \\
x_5[t] &= 1.5x_4[t-1] + 0.6x_6[t-1] + e_5[t] \\
x_6[t] &= 0.7x_5[t-1] + 0.5x_7[t-1] + e_6[t] \\
x_7[t] &= 0.65x_6[t-1] + e_7[t]
\end{aligned}
\tag{7.7}
$$

where e_i are white noise sequences. The corruption model for node 4 is:

$$u_4[t] = \begin{cases} x_4[t-2], & \text{with probability } 0.7 \\ x_4[t], & \text{with probability } 0.3. \end{cases} \tag{7.8}$$

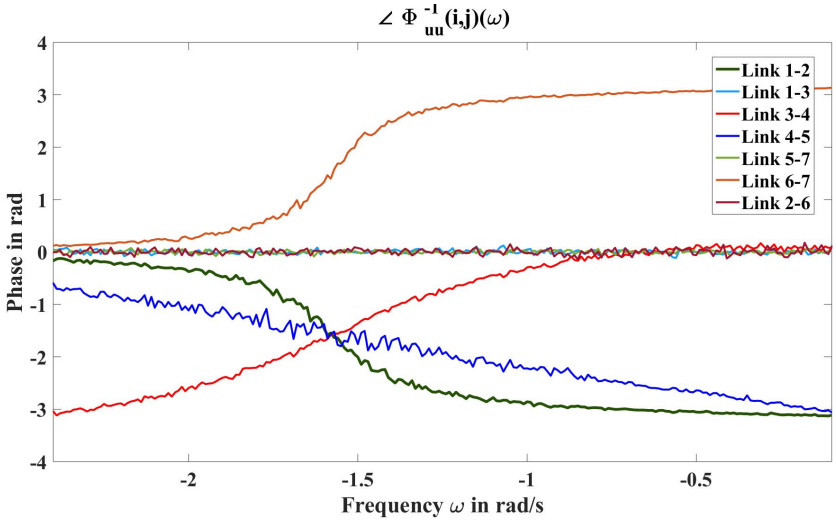


Figure 7.4: **Phase Plots**. The estimated phase response values are shown in the figure. We observe that the phase response corresponding to edges to the leaf nodes, $\{1, 7\}$, have only non-constant phase response. Node 4 has two neighbors with non-constant phase response. This verifies predictions of Theorem 10. The phase response estimate of $2-6$ link is approximately close to zero and is a constant. This verifies Lemma 5.

From a trajectory length of 10^7, the estimates for power spectral density was obtained using MATLAB 'cpsd' command. The plot for magnitude of the inverse power spectral density estimates is shown in Figure 7.3. *Step 1:* Using Theorem 3, adding edges and constructing an undirected graph results in the perturbed graph shown in figure 7.2(c)). *Step 2:* We notice that neighbors of 1,4 and 7 forms a clique with nodes 1,4 and 7 respectively. As predicted by Lemma 2, we have identified the candidate set, $\{1, 4, 7\}$, containing corrupt nodes and leaf nodes only. *Step 3:* The next step is to detect the corrupt node. To this we observe the phase response of the inverse PSD estimated. Figure 7.4 shows that only 4 will have at least two non-constant phase estimates. For leaf nodes, there will only be one non-constant phase plot. Using Theorem

10 we determine node 4 as the corrupt node. *Step 5:* The next step is to follow the *hide and learn* algorithm. We first remove the measurements of node 4 and infer the topology of the network with latent node 4. That is, using the measurements $O = \{1, 2, 3, 5, 6, 7\}$, we compute the inverse PSD. The magnitude of Φ_{oo}^{-1} is shown in Figure 7.5. Following Lemma 3 yields the undirected graph shown in figure 7.6(a)).

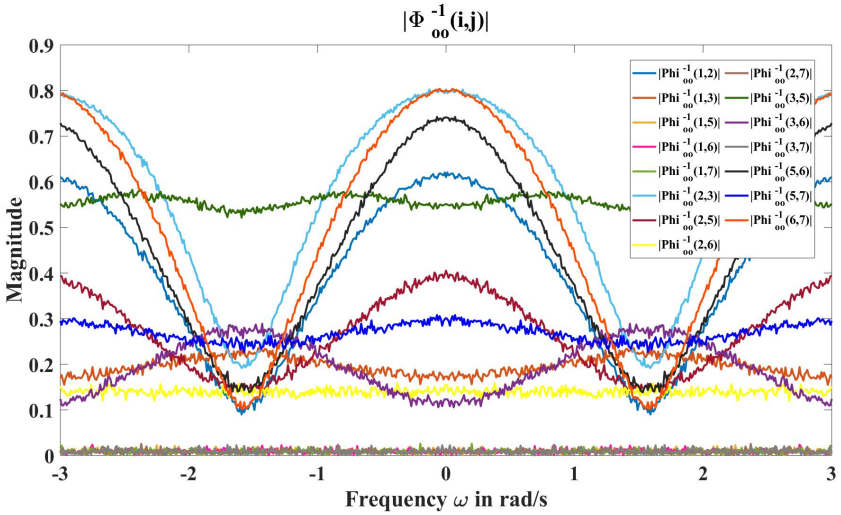
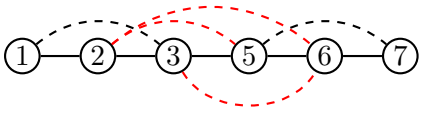
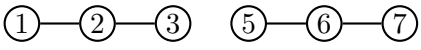


Figure 7.5: **Magnitude plots with unobserved node** 4. The magnitude of the inverse PSD estimates computed from $o = u \setminus \{4\}$ are shown here. y axis is angular frequency ω in radians/s. Notice the entries are non-zero across the frequency grid. To each non-zero entry, we add undirected edges to infer the undirected graph as shown in Figure 7.6(a)) following Lemma 3.



(a) Inferred undirected graph with latent node 4.



(b) Detect true edges

(c) Place hidden (corrupt) node 4.

Figure 7.6: Illustration of Hide and Learn algorithm.

Then, using graphical separation results mentioned in Lemma 4, we detect the true edges in the inferred network. This yields two disconnected components as shown in Figure 7.6(b)). Finally, we place the latent node at the point of disconnection and obtain the true generative topology shown in Figure 7.6(c)).

7.5 Summary

In this chapter, we proposed an exact topology learning algorithm for radial bi-directed network of LTI systems in the presence of corruption. We show that for networks where corrupt nodes are three or more hops away from each other deep inside the network, the spurious edges owning to data corruption can be eliminated and the the exact network structure can be determined. We used clique characterization in the inferred undirected graph to determine the set of leaf and corrupt nodes. Then using phase properties of the inverse PSD, we isolated the location of corrupt nodes. Finally, we hide the corrupt node measurements and adopt *hide and learn* strategy to learn the exact network representation generating the time series observations. We remark here that Algorithm 2 and 3 will still work to learn the exact network structure even when there are hidden nodes and corruption simultaneously as long as the location of hidden nodes and the corrupt nodes are at least 3 hops away from each other and at least 3 hops away from the leaf nodes.

Chapter 8

Future Directions

The primary emphasis of this research was to incorporate realistic modeling assumptions on data streams and characterize the effects of data corruption on network identification using passive means. We showed that identifying the structure of networked systems using corrupt measurements results in the inference of spurious links. Moreover, we provided exact characterization on the location of such spurious links. Application of our results for network design and fault/attack-detection could be explored in specific cyber-physical application domains such as heat-transfer models and power distribution grids. Such experimental analysis will significantly broaden the scope of techniques pursued in this dissertation into cyber-security domain.

A pertinent direction of future work could be on quantifying the amount of data that is needed to detect network inter-relationships using statistical measures of dependency. In this dissertation, we used conditional directed information (DI) measure and studied the asymptotic consistency of DI measures focusing on finite alphabets. In many practical applications, the amount of measurements available are finite and continuous valued. Thus, it is necessary to quantify the amount of samples required and provide confidence intervals in estimating DI from finite samples of data that are not restricted to finite alphabets.

An interesting problem to work in future could be to quantitatively differentiate between the true and spurious influences. It will be interesting to devise sufficient statistical conditions on computed DI that delineates these differences in the estimated DI due to data corruption. This could entail rigorously devising sufficient statistical

conditions on corruption models and quantifying the effects on measures of statistical dependency in networked time-series. Such assessment can facilitate elimination of spurious links that arise due to measurement uncertainties.

Our results serve as a necessary first step in understanding what part of network reconstruction can be trusted to facilitate accurate system identification. We can begin with linear time-invariant systems and generalize to consider nonlinear dynamical systems. Verifying identifiability conditions [82], [103] and quantifying error in the identified transfer function due to data corruption from true dynamics can be another interesting line of work. Another interesting direction could be to consider network reconstruction for non-target specific nonlinear dynamical systems. Non-target specific network reconstruction for linear systems studied in [192] and [79] may yield useful insights in this direction.

We considered the problem of detecting corrupt nodes and exact topology inference for *radial* bi-directed systems. Extension entails relaxing the assumption that the topology is a tree, and consider generic graphical structures that underlie the networked time-series. Inclusion of correlated disturbances and latent nodes framework will further expand the scope of network identification.

References

[1] Jingfang Fan, Jun Meng, Yosef Ashkenazy, Shlomo Havlin, and Hans Joachim Schellnhuber. Network analysis reveals strongly localized impacts of el niño. *Proceedings of the National Academy of Sciences*, 2017.

[2] Henk A Dijkstra, Emilio Hernández-García, Cristina Masoller, and Marcelo Barreiro. *Networks in climate*. Cambridge University Press, 2019.

[3] Adam Vaccaro, Julien Emile-Geay, Dominque Guillot, Resherle Verna, Colin Morice, John Kennedy, and Bala Rajaratnam. Climate field completion via markov random fields–application to the hadcrut4. 6 temperature dataset. *Journal of Climate*, pages 1–66, 2021.

[4] Nicholas C Grassly and Christophe Fraser. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology*, 6(6):477–487, 2008.

[5] Jay S Kaufman. *Methods in social epidemiology*, volume 16. John Wiley & Sons, 2017.

[6] Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, 20(3):353, 2017.

[7] Stefano Anzellotti, Dorit Kliemann, Nir Jacoby, and Rebecca Saxe. Directed network discovery with dynamic network modelling. *Neuropsychologia*, 99:1–11, 2017.

[8] Stefan Frässle, Ekaterina I Lomakina, Lars Kasper, Zina M Manjaly, Alex Leff, Klaas P Pruessmann, Joachim M Buhmann, and Klaas E Stephan. A generative model of whole-brain effective connectivity. *Neuroimage*, 179:505–529, 2018.

[9] A Agung Julius and Calin Belta. Genetic regulatory network identification using monotone functions decomposition. *IFAC Proceedings Volumes*, 44(1):11785–11790, 2011.

[10] Yiming Zuo, Yi Cui, Guoqiang Yu, Ruijiang Li, and Habtom W Ressom. Incorporating prior biological knowledge for network-based differential gene expression analysis using differentially weighted graphical lasso. *BMC bioinformatics*, 18(1):1–14, 2017.

[11] Mariano Rubiolo, Diego H Milone, and Georgina Stegmayer. Extreme learning machines for reverse engineering of gene regulatory networks from expression time series. *Bioinformatics*, 34(7):1253–1260, 2018.

[12] Vasco M Carvalho. From micro to macro via production networks. *Journal of Economic Perspectives*, 28(4):23–48, 2014.

[13] Paolo Giudici and Alessandro Spelta. Graphical network models for international financial flows. *Journal of Business & Economic Statistics*, 34(1):128–138, 2016.

[14] Dror Y Kenett and Shlomo Havlin. Network science: a useful tool in economics and finance. *Mind & Society*, 14(2):155–167, 2015.

[15] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. Context aware computing for the internet of things: A survey. *IEEE communications surveys & tutorials*, 16(1):414–454, 2014.

[16] D. Guinard, V. Trifa, S. Karnouskos, P. Spiess, and D. Savio. Interacting with the soa-based internet of things: Discovery, query, selection, and on-demand provisioning of web services. *IEEE Transactions on Services Computing*, 3(3):223–235, 2010.

[17] Mohammad Saeid Mahdavinejad, Mohammadreza Rezvan, Mohammadamin Barekatain, Peyman Adibi, Payam Barnaghi, and Amit P Sheth. Machine learning for internet of things data analysis: A survey. *Digital Communications and Networks*, 4(3):161–175, 2018.

[18] Yang Wang, Tele Tan, and Kia-Fock Loe. Video segmentation based on graphical models. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–335. IEEE, 2003.

[19] Ian F Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2002.

[20] Georgios B Giannakis, Vassilis Kekatos, Nikolaos Gatsis, Seung-Jun Kim, Hao Zhu, and Bruce F Wollenberg. Monitoring and optimization for power grids: A signal processing perspective. *IEEE Signal Processing Magazine*, 30(5):107–128, 2013.

[21] D. Deka, S. Backhaus, and M. Chertkov. Structure learning in power distribution networks. *IEEE Transactions on Control of Network Systems*, 5(3):1061–1074, Sept 2018.

[22] John Scott. Social network analysis. *Sociology*, 22(1):109–127, 1988.

[23] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6):065701, 2011.

[24] Xiang Xu, Cheng Zhu, Qingyong Wang, Xianqiang Zhu, and Yun Zhou. Identifying vital nodes in complex networks by adjacency information entropy. *Scientific reports*, 10(1):1–12, 2020.

[25] Reza Olfati-Saber and Richard M Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on automatic control*, 49(9):1520–1533, 2004.

[26] O Patrick Kreidl. *Graphical models and message-passing algorithms for network-constrained decision problems.* PhD thesis, Massachusetts Institute of Technology, 2008.

[27] Wei Wang, Quan-Hui Liu, Junhao Liang, Yanqing Hu, and Tao Zhou. Coevolution spreading in complex networks. *Physics Reports*, 820:1–51, 2019.

[28] Ye Yuan, Xiuchuan Tang, Wei Zhou, Wei Pan, Xiuting Li, Hai-Tao Zhang, Han Ding, and Jorge Goncalves. Data driven discovery of cyber physical systems. *Nature communications*, 10(1):1–9, 2019.

[29] Leonidas Georgiadis, Michael J Neely, and Leandros Tassiulas. *Resource allocation and cross-layer control in wireless networks*. Now Publishers Inc, 2006.

[30] Ali Dorri, Salil S Kanhere, and Raja Jurdak. Multi-agent systems: A survey. *Ieee Access*, 6:28573–28593, 2018.

[31] Harald Sundmaeker, Patrick Guillemin, Peter Friess, and Sylvie Woelfflé. Vision and challenges for realising the internet of things. *Cluster of European Research Projects on the Internet of Things, European Commision*, 2010.

[32] D. Datla, X. Chen, T. Tsou, S. Raghunandan, S. Hasan, S., H. Reed, J., B. Dietrich, C., T. Bose, B. Fette, and H. Kim, J. Wireless distributed computing: a survey of research challenges. *IEEE Communications Magazine*, 50(1):144–152, 2012.

[33] Yang-Bo He and Zhi Geng. Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9(Nov):2523–2547, 2008.

[34] Wray Buntine. A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on knowledge and data engineering*, 8(2):195–210, 1996.

[35] Marzieh Nabi-Abdolyousefi and Mehran Mesbahi. Network identification via node knockout. *IEEE Transactions on Automatic Control*, 57(12):3214–3219, 2012.

[36] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[37] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[38] Jiming Chen, Karl H Johansson, Stephan Olariu, Ioannis Ch Paschalidis, and Ivan Stojmenovic. Guest editorial special issue on wireless sensor and actuator networks. *IEEE Transactions on Automatic Control*, 56(10):2244–2246, 2011.

[39] Man Zhang, Bran Selic, Shaukat Ali, Tao Yue, Oscar Okariz, and Roland Norgren. Understanding uncertainty in cyber-physical systems: a conceptual model. In *European conference on modelling foundations and applications*, pages 247–264. Springer, 2016.

[40] Deepak Ganesan, Sylvia Ratnasamy, Hanbiao Wang, and Deborah Estrin. Coping with irregular spatio-temporal sampling in sensor networks. *ACM SIGCOMM Computer Communication Review*, 34(1):125–130, 2004.

[41] Hsin-Hung Cho, Chi-Yuan Chen, Timothy K Shih, and Han-Chieh Chao. Survey on underwater delay/disruption tolerant wireless sensor network routing. *IET Wireless Sensor Systems*, 4(3):112–121, 2014.

[42] Lu Jiang, Liping Yan, Yuanqing Xia, Qiao Guo, Mengyin Fu, and Kunfeng Lu. Asynchronous multirate multisensor data fusion over unreliable measurements with correlated noise. *IEEE Transactions on Aerospace and Electronic Systems*, 53(5):2427–2437, 2017.

[43] Shuli Sun, Fangfang Peng, and Honglei Lin. Distributed asynchronous fusion estimator for stochastic uncertain systems with multiple sensors of different fading measurement rates. *IEEE Transactions on Signal Processing*, 66(3):641–653, 2017.

[44] Xingzhen Bai, Zidong Wang, Li Sheng, and Zhen Wang. Reliable data fusion of hierarchical wireless sensor networks with asynchronous measurement for greenhouse monitoring. *IEEE Transactions on Control Systems Technology*, 27(3):1036–1046, 2018.

[45] Guido Cavraro, Emiliano Dall'Anese, and Andrey Bernstein. Dynamic power network state estimation with asynchronous measurements. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2019.

[46] Vicent Girbes-Juan, Leopoldo Armesto, Daniel Hernandez-Ferrandiz, Juan Francisco Dols, and Antonio Sala. Asynchronous sensor fusion of gps, imu and can-based odometry for heavy-duty vehicles. *IEEE Transactions on Vehicular Technology*, 2021.

[47] Dae-Hyun Choi and Le Xie. Sensitivity analysis of real-time locational marginal price to scada sensor data corruption. *IEEE Transactions on Power Systems*, 29(3):1110–1120, 2013.

[48] Zheng Yang, Lirong Jian, Chenshu Wu, and Yunhao Liu. Beyond triangle inequality: Sifting noisy and outlier distance measurements for localization. *ACM Transactions on Sensor Networks (TOSN)*, 9(2):1–20, 2013.

[49] Milos S. Stankovic, Srdjan S. Stankovic, and Karl Henrik Johansson. Distributed time synchronization for networks with random delays and measurement noise. *Automatica*, 93:126 – 137, 2018.

[50] Hong-Ning Dai, Raymond Chi-Wing Wong, Hao Wang, Zibin Zheng, and Athanasios V Vasilakos. Big data analytics for large-scale wireless networks: Challenges and opportunities. *ACM Computing Surveys (CSUR)*, 52(5):1–36, 2019.

[51] Manolis Tsakiris and Liangzu Peng. Homomorphic sensing. In *International Conference on Machine Learning*, pages 6335–6344. PMLR, 2019.

[52] Štefan Knotek, Kristian Hengster-Movric, and Michael Šebek. Distributed estimation on sensor networks with measurement uncertainties. *IEEE Transactions on Control Systems Technology*, 2020.

[53] Daniel E Quevedo and Isabel Jurado. Stability of sequence-based control with random delays and dropouts. *IEEE Transactions on Automatic Control*, 59(5):1296–1302, 2013.

[54] A. S. Leong, S. Dey, and D. E. Quevedo. Sensor scheduling in variance based event triggered estimation with packet drops. *IEEE Transactions on Automatic Control*, 62(4):1880–1895, 2017.

[55] Prabhat K Mishra, Debasish Chatterjee, and Daniel E Quevedo. Stabilizing stochastic predictive control under bernoulli dropouts. *IEEE Transactions on Automatic Control*, 63(6):1579–1590, 2017.

[56] Jianming Zhou, Guoxiang Gu, and Xiang Chen. Distributed kalman filtering over wireless sensor networks in the presence of data packet drops. *IEEE Transactions on Automatic Control*, 64(4):1603–1610, 2018.

[57] Jiangkai Peng, Bo Fan, Hao Xu, and Wenxin Liu. Discrete-time self-triggered control of dc microgrids with data dropouts and communication delays. *IEEE Transactions on Smart Grid*, 11(6):4626–4636, 2020.

[58] Alvaro Cardenas, Saurabh Amin, Bruno Sinopoli, Annarita Giani, Adrian Perrig, Shankar Sastry, et al. Challenges for securing cyber physical systems. In *Workshop on future directions in cyber-physical systems security*, volume 5. Citeseer, 2009.

[59] Yuzhe Li, Ling Shi, Peng Cheng, Jiming Chen, and Daniel E Quevedo. Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach. *IEEE Transactions on Automatic Control*, 60(10):2831–2836, 2015.

[60] Lifeng Ma, Zidong Wang, Qing-Long Han, and Hak-Keung Lam. Variance-constrained distributed filtering for time-varying systems with multiplicative noises and deception attacks over sensor networks. *IEEE Sensors Journal*, 17(7):2279–2288, 2017.

[61] Ruilong Deng, Peng Zhuang, and Hao Liang. False data injection attacks against state estimation in power distribution systems. *IEEE Transactions on Smart Grid*, 10(3):2871–2881, 2018.

[62] Tien Nguyen, Shiyuan Wang, Mohannad Alhazmi, Mostafa Nazemi, Abouzar Estebsari, and Payman Dehghanian. Electric power grid resilience to cyber adversaries: State of the art. *IEEE Access*, 8:87592–87608, 2020.

[63] Andrew Bolstad, Barry D Van Veen, and Robert Nowak. Causal network inference via group sparse regularization. *IEEE transactions on signal processing*, 59(6):2628–2641, 2011.

[64] Alessandro Chiuso and Gianluigi Pillonetto. A bayesian approach to sparse dynamic network identification. *Automatica*, 48(8):1553–1565, 2012.

[65] Yin Wang, Mario Sznaier, and O Camps. A super-atomic norm minimization approach to identifying sparse dynamical graphical models. In *2016 American Control Conference (ACC)*, pages 1962–1967. IEEE, 2016.

[66] David Hayden, Young Hwan Chang, Jorge Goncalves, and Claire J Tomlin. Sparse network identifiability via compressed sensing. *Automatica*, 68:9–17, 2016.

[67] Mor Nitzan, Jose Casadiego, and Marc Timme. Revealing physical interaction networks from statistics of collective dynamics. *Science advances*, 3(2):e1600396, 2017.

[68] Santiago Segarra, Antonio G Marques, Gonzalo Mateos, and Alejandro Ribeiro. Network topology inference from spectral templates. *IEEE Transactions on Signal and Information Processing over Networks*, 3(3):467–483, 2017.

[69] Arun Venkitaraman, Håkan Hjalmarsson, and Bo Wahlberg. Learning sparse linear dynamic networks in a hyper-parameter free setting. *IFAC-PapersOnLine*, 53(2):75–79, 2020.

[70] Shahin Shahrampour and Victor M Preciado. Topology identification of directed dynamical networks via power spectral analysis. *IEEE Transactions on Automatic Control*, 60(8):2260–2265, 2014.

[71] Mario Coutino, Elvin Isufi, Takanori Maehara, and Geert Leus. State-space network topology identification from partial observations. *IEEE Transactions on Signal and Information Processing over Networks*, 6:211–225, 2020.

[72] Xiufeng Zhang, Gang Wang, Tao Cai, and Jian Sun. Network topology identification under the multi-agent agreement protocol. *Journal of the Franklin Institute*, 2021.

[73] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

[74] Donatello Materassi and Murti V Salapaka. On the problem of reconstructing an unknown topology via locality properties of the wiener filter. *IEEE transactions on automatic control*, 57(7):1765–1777, 2012.

[75] Arne G Dankers, Paul MJ Van den Hof, Peter SC Heuberger, and Xavier Bombois. Dynamic network structure identification with prediction error methods-basic examples. *IFAC Proceedings Volumes*, 45(16):876–881, 2012.

[76] David Hayden, Ye Yuan, and Jorge Goncalves. Network reconstruction from intrinsic noise: Non-minimum-phase systems. *IFAC Proceedings Volumes*, 47(3), 2014.

[77] Lionel Barnett and Anil K Seth. Granger causality for state-space models. *Physical Review E*, 91(4):040101, 2015.

[78] C. J. Quinn, N. Kiyavash, and T. P. Coleman. Directed Information Graphs. *IEEE Transactions on Information Theory*, 61(12):6887–6909, 2015.

[79] Vasu Chetty, Joel Eliason, and Sean Warnick. Passive reconstruction of non-target-specific discrete-time lti systems. In *American Control Conference (ACC)*. IEEE, 2016.

[80] Jalal Etesami, Negar Kiyavash, and Todd Coleman. Learning minimal latent directed information polytrees. *Neural computation*, 28(9):1723–1768, 2016.

[81] Harm Weerts, Paul MJ Van den Hof, and Arne Dankers. Identification of dynamic networks operating in the presence of algebraic loops. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 4606–4611. IEEE, 2016.

[82] Harm H.M. Weerts, Paul M.J. Van den Hof, and Arne G. Dankers. Identifiability of linear dynamic networks. *Automatica*, 89:247 – 258, 2018.

[83] Giulio Bottegal, Alessandro Chiuso, and Paul MJ Van den Hof. On dynamic network modeling of stationary multivariate processes. *IFAC-PapersOnLine*, 51(15):850–855, 2018.

[84] Henk J van Waarde, Pietro Tesi, and M Kanat Camlibel. Topology reconstruction of dynamical networks via constrained lyapunov equations. *IEEE Transactions on Automatic Control*, 64(10):4300–4306, 2019.

[85] Firoozeh Sepehr and Donatello Materassi. Blind learning of tree network topologies in the presence of hidden nodes. *IEEE Transactions on Automatic Control*, 2019.

[86] Saurav Talukdar, Deepjyoti Deka, Harish Doddi, Donatello Materassi, Michael Chertkov, and Murti V Salapaka. Physics informed topology learning in networks of linear dynamical systems. *Automatica*, 112:108705, 2020.

[87] Firoozeh Sepehr and Donatello Materassi. Non-invasive approximation of linear dynamic system networks using polytrees. *IEEE Transactions on Control of Network Systems*, 2021.

[88] Yanning Shen, Xiao Fu, Georgios B Giannakis, and Nicholas D Sidiropoulos. Topology identification of directed graphs via joint diagonalization of correlation matrices. *IEEE Transactions on Signal and Information Processing over Networks*, 6:271–283, 2020.

[89] Mihaela Dimovska and Donatello Materassi. Granger-faithfulness and link orientation in network reconstruction. *arXiv preprint arXiv:2012.02332*, 2020.

[90] J. Goncalves and S. Warnick. Necessary and sufficient conditions for dynamical structure reconstruction of lti networks. *IEEE Transactions on Automatic Control*, 53(7):1670–1674, Aug 2008.

[91] Alexandre S Bazanella, Michel Gevers, Julien M Hendrickx, and Adriane Parraga. Identifiability of dynamical networks: which nodes need be measured? In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 5870–5875. IEEE, 2017.

[92] Xiaodong Cheng, Shengling Shi, and Paul MJ Van den Hof. Allocation of excitation signals for generic identifiability of dynamic networks. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 5507–5512. IEEE, 2019.

[93] Henk J van Waarde, Pietro Tesi, and M Kanat Camlibel. Topology identification of heterogeneous networks: Identifiability and reconstruction. *Automatica*, 123:109331, 2021.

[94] Ananth Narayan Samudrala, M Hadi Amini, Soummya Kar, and Rick S Blum. Sensor placement for outage identifiability in power distribution networks. *IEEE Transactions on Smart Grid*, 11(3):1996–2013, 2019.

[95] C Giudicianni, M Herrera, A Di Nardo, R Greco, E Creaco, and A Scala. Topological placement of quality sensors in water-distribution networks without the recourse to hydraulic modeling. *Journal of Water Resources Planning and Management*, 146(6):04020030, 2020.

[96] C Arcadius Tokognon, Bin Gao, Gui Yun Tian, and Yan Yan. Structural health monitoring framework based on internet of things: A survey. *IEEE Internet of Things Journal*, 4(3):619–635, 2017.

[97] Carmelo Scuro, Paolo Francesco Sciammarella, Francesco Lamonaca, Renato Sante Olivito, and Domenico Luca Carni. Iot for structural health monitoring. *IEEE Instrumentation & Measurement Magazine*, 21(6):4–14, 2018.

[98] Andrew Blake, Pushmeet Kohli, and Carsten Rother. *Markov random fields for vision and image processing*. Mit Press, 2011.

[99] Siyi Hu and Luca Carlone. Accelerated inference in markov random fields via smooth riemannian optimization. *IEEE Robotics and Automation Letters*, 4(2):1295–1302, 2019.

[100] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

[101] Martin J Wainwright and Michael Irwin Jordan. *Graphical models, exponential families, and variational inference*. Now Publishers Inc, 2008.

[102] Mathias Drton and Marloes H Maathuis. Structure learning in graphical modeling. *Annual Review of Statistics and Its Application*, 4:365–393, 2017.

[103] J. M. Hendrickx, M. Gevers, and A. S. Bazanella. Identifiability of dynamical networks with partial node measurements. *IEEE Transactions on Automatic Control*, 2018.

[104] Niklas Everitt, Giulio Bottegal, and Håkan Hjalmarsson. An empirical bayes approach to identification of modules in dynamic networks. *Automatica*, 91:144 – 151, 2018.

[105] Henk J Van Waarde, Pietro Tesi, and M Kanat Camlibel. Necessary and sufficient topological conditions for identifiability of dynamical networks. *IEEE Transactions on Automatic Control*, 2019.

[106] Arne Dankers, Paul MJ Van den Hof, Xavier Bombois, and Peter SC Heuberger. Identification of dynamic models in complex networks with prediction error methods: Predictor input selection. *IEEE Transactions on Automatic Control*, 61(4):937–952, 2015.

[107] Chengpu Yu and Michel Verhaegen. Subspace identification of distributed clusters of homogeneous systems. *IEEE Transactions on Automatic Control*, 62(1):463–468, 2016.

[108] Michel Gevers, Alexandre S. Bazanella, and Adriane Parraga. On the identifiability of dynamical networks. *IFAC-PapersOnLine*, 50(1), 2017. 20th IFAC World Congress.

[109] Michel Gevers, Alexandre Sanfelice Bazanella, and Gian Vianna da Silva. A practical method for the consistent identification of a module in a dynamical network. *IFAC-PapersOnLine*, 51(15):862–867, 2018.

[110] Karthik R Ramaswamy, Giulio Bottegal, and Paul MJ Van den Hof. Local module identification in dynamic networks using regularized kernel-based methods. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 4713–4718. IEEE, 2018.

[111] Harm HM Weerts, Paul MJ Van den Hof, and Arne G Dankers. Prediction error identification of linear dynamic networks with rank-reduced noise. *Automatica*, 98:256–268, 2018.

[112] Donatello Materassi and Murti V Salapaka. Signal selection for estimation and identification in networks of dynamic systems: a graphical model approach. *IEEE Transactions on Automatic Control*, 65(10):4138–4153, 2019.

[113] Chengpu Yu, Jie Chen, and Michel Verhaegen. Subspace identification of individual systems in a large-scale heterogeneous network. *Automatica*, 109:108517, 2019.

[114] K. R. Ramaswamy and P. M. J. Vandenhof. A local direct method for module identification in dynamic networks with correlated noise. *IEEE Transactions on Automatic Control*, 2020.

[115] Saverio Bolognani, Nicoletta Bof, Davide Michelotti, Riccardo Muraro, and Luca Schenato. Identification of power distribution network topology via voltage correlation analysis. In *52nd IEEE Conference on Decision and Control*, pages 1659–1664. IEEE, 2013.

[116] Saurav Talukdar, Deepjyoti Deka, Blake Lundstrom, Michael Chertkov, and Murti V Salapaka. Learning exact topology of a loopy power grid from ambient dynamics. In *Proceedings of the Eighth International Conference on Future Energy Systems*, pages 222–227, 2017.

[117] Deepjyoti Deka, Michael Chertkov, and Scott Backhaus. Topology estimation using graphical models in multi-phase power distribution grids. *IEEE Transactions on Power Systems*, 35(3):1663–1673, 2019.

[118] Kunihiko Taira, Aditya G Nair, and Steven L Brunton. Network structure of two-dimensional decaying isotropic turbulence. *Journal of Fluid Mechanics*, 795, 2016.

[119] Alexandros Alexakis and Luca Biferale. Cascades and transitions in turbulent flows. *Physics Reports*, 767:1–101, 2018.

[120] Jin-Han Xie and Oliver Bühler. Third-order structure functions for isotropic turbulence with bidirectional energy transfer. *Journal of Fluid Mechanics*, 877, 2019.

[121] Zhe Bai, N Benjamin Erichson, Muralikrishnan Gopalakrishnan Meena, Kunihiko Taira, and Steven L Brunton. Randomized methods to characterize large-scale vortical flow networks. *PloS one*, 14(11):e0225265, 2019.

[122] R Zarin Pass, Michael Wetter, and Mary Ann Piette. A thermodynamic analysis of a novel bidirectional district heating and cooling network. *Energy*, 144:20–30, 2018.

[123] Justus von Rhein, Gregor P Henze, Nicholas Long, and Yangyang Fu. Development of a topology analysis tool for fifth-generation district heating and cooling networks. *Energy conversion and management*, 196:705–716, 2019.

[124] Marco Wirtz, Lukas Kivilip, Peter Remmen, and Dirk Müller. Quantifying demand balancing in bidirectional low temperature networks. *Energy and Buildings*, 224:110245, 2020.

[125] Tobias Sommer, Matthias Sulzer, Michael Wetter, Artem Sotnikov, Stefan Mennel, and Christoph Stettler. The reservoir network: A new network topology for district heating and cooling. *Energy*, 199:117418, 2020.

[126] Mattia Zorzi and Rodolphe Sepulchre. Ar identification of latent-variable graphical models. *IEEE Transactions on Automatic Control*, 61(9):2327–2340, 2015.

[127] Saurav Talukdar, Deepjyoti Deka, Michael Chertkov, and Murti Salapaka. Topology learning of radial dynamical systems with latent nodes. In *2018 Annual American Control Conference (ACC)*, pages 1096–1101. IEEE, 2018.

[128] EMM Lizan Kivits and Paul MJ Van den Hof. A dynamic network approach to identification of physical systems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 4533–4538. IEEE, 2019.

[129] Arne Dankers, Paul MJ Van den Hof, Xavier Bombois, and Peter SC Heuberger. Errors-in-variables identification in dynamic networks—consistency results for an instrumental variable approach. *Automatica*, 62, 2015.

[130] Patricio Torres, Jan-Willem van Wingerden, and Michel Verhaegen. Po-moesp subspace identification of directed acyclic graphs with unknown topology. *Automatica*, 53:60–71, 2015.

[131] Jonas Linder and Martin Enqvist. Identification of systems with unknown inputs using indirect input measurements. *International Journal of Control*, 90(4):729–745, 2017.

[132] Torsten Söderström. *Errors-in-variables methods in system identification*. Springer, 2018.

[133] Y. Yuan, G. B. Stan, S. Warnick, and J. Goncalves. Robust dynamical network structure reconstruction. *Automatica*, 47(6):1230 – 1235, 2011. Special Issue on Systems Biology.

[134] V. Chetty, D. Hayden, J. Goncalves, and S. Warnick. Robust signal-structure reconstruction. In *52nd IEEE Conference on Decision and Control*, pages 3184–3189, Dec 2013.

[135] J. Goncalves and S. Warnick. Necessary and sufficient conditions for dynamical structure reconstruction of lti networks. *IEEE Transactions on Automatic Control*, 53(7):1670–1674, Aug 2008.

[136] David Hayden, Ye Yuan, and Jorge Gonçalves. Network reconstruction from intrinsic noise: Minimum-phase systems. In *2014 American Control Conference*. IEEE, 2014.

[137] K. B. Petersen and M. S. Pedersen. The matrix cookbook, 2012.

[138] Michael Banf and Seung Y Rhee. Enhancing gene regulatory network inference through data integration with markov random fields. *Scientific reports*, 7(1):1–13, 2017.

[139] Jinyuan Jia, Binghui Wang, Le Zhang, and Neil Zhenqiang Gong. Attriinfer: Inferring user attributes in online social networks using markov random fields. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1561–1569, 2017.

[140] Guy Bresler, David Gamarnik, and Devavrat Shah. Structure learning of antiferromagnetic ising models. In *Advances in Neural Information Processing Systems*, pages 2852–2860, 2014.

[141] Chuan Li and Michael Wand. Combining markov random fields and convolutional neural networks for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2479–2486, 2016.

[142] Baoyuan Wu, Bao-Gang Hu, and Qiang Ji. A coupled hidden markov random field model for simultaneous face clustering and tracking in videos. *Pattern Recognition*, 64:361–373, 2017.

[143] Linh Nguyen, Sarath Kodagoda, Ravindra Ranasinghe, and Gamini Dissanayake. Mobile robotic sensors for environmental monitoring using gaussian markov random field. *Robotica*, 39(5):862–884, 2021.

[144] Tetsuya Takaishi. Multiple time series ising model for financial market simulations. In *Journal of Physics: Conference Series*, volume 574, page 012149. IOP Publishing, 2015.

[145] Tatsu Kuwatani, Kenji Nagata, Masato Okada, and Mitsuhiro Toriumi. Markov random field modeling for mapping geofluid distributions from seismic velocity structures. *Earth, Planets and Space*, 66(1):1–9, 2014.

[146] Brian J Reich and Benjamin A Shaby. A spatial markov model for climate extremes. *Journal of Computational and Graphical Statistics*, 28(1):117–126, 2019.

[147] Michael Irwin Jordan. *Learning in graphical models*, volume 89. Springer Science & Business Media, 1998.

[148] Martin J Wainwright and Michael I Jordan. Log-determinant relaxation for approximate inference in discrete markov random fields. *IEEE transactions on signal processing*, 54(6):2099–2109, 2006.

[149] Su-In Lee, Varun Ganapathi, and Daphne Koller. Efficient structure learning of markov networks using l 1-regularization. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, pages 817–824, 2006.

[150] Bernhard Schölkopf, John Platt, and Thomas Hofmann. Efficient structure learning of markov networks using l1-regularization. 2007.

[151] Holger Höfling and Robert Tibshirani. Estimation of sparse binary pairwise markov networks using pseudo-likelihoods. *Journal of Machine Learning Research*, 10(4), 2009.

[152] Guy Bresler, Elchanan Mossel, and Allan Sly. Reconstruction of markov random fields from samples: Some observations and algorithms. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 343–356. Springer, 2008.

[153] Pradeep Ravikumar, Martin J Wainwright, and John D Lafferty. High-dimensional ising model selection using l1-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319, 2010.

[154] Avik Ray, Sujay Sanghavi, and Sanjay Shakkottai. Improved greedy algorithms for learning graphical models. *IEEE Transactions on Information Theory*, 61(6):3457–3468, 2015.

[155] Youngsuk Park, David Hallac, Stephen Boyd, and Jure Leskovec. Learning the network structure of heterogeneous data via pairwise exponential markov random fields. In *Artificial Intelligence and Statistics*, pages 1302–1310. PMLR, 2017.

[156] Linus Hamilton, Frederic Koehler, and Ankur Moitra. Information theoretic properties of markov random fields, and their algorithmic applications. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 2460–2469, 2017.

[157] Shanshan Wu, Sujay Sanghavi, and Alexandros G Dimakis. Sparse logistic regression learns all discrete pairwise graphical models. *Advances in Neural Information Processing Systems*, 32:8071–8081, 2019.

[158] Po-Ling Loh and Martin J Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, 40(3):1637–1664, 2012.

[159] Mladen Kolar and Eric P Xing. Estimating sparse precision matrices from data with missing values. In *International Conference on Machine Learning*, pages 635–642, 2012.

[160] Yudong Chen and Constantine Caramanis. Noisy and missing data regression: Distribution-oblivious support recovery. In *International Conference on Machine Learning*, pages 383–391. PMLR, 2013.

[161] Eunho Yang and Aurélie C Lozano. Robust gaussian graphical modeling with the trimmed graphical lasso. In *Advances in Neural Information Processing Systems*, pages 2602–2610, 2015.

[162] Lingxiao Wang and Quanquan Gu. Robust gaussian graphical model estimation with arbitrary corruption. In *International Conference on Machine Learning*, pages 3617–3626. PMLR, 2017.

[163] Yu Cheng, Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Robust learning of fixed-structure bayesian networks. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018*, 2018.

[164] Erik M Lindgren, Vatsal Shah, Yanyao Shen, Alexandros G Dimakis, and Adam Klivans. On robust learning of ising models. In *NeurIPS Workshop on Relational Representation Learning*, 2019.

[165] Adarsh Prasad, Vishwak Srinivasan, Sivaraman Balakrishnan, and Pradeep Ravikumar. On learning ising models under huber's contamination model. *Advances in neural information processing systems*, 33, 2020.

[166] Konstantinos E Nikolakakis, Dionysios S Kalogerias, and Anand D Sarwate. Learning tree structures from noisy data. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1771–1782. PMLR, 2019.

[167] Surbhi Goel, Daniel M Kane, and Adam R Klivans. Learning ising models with independent failures. In *Conference on Learning Theory*, pages 1449–1469. PMLR, 2019.

[168] Fengzhuo Zhang and Vincent YF Tan. Robustifying algorithms of learning latent trees with vector variables. *arXiv preprint arXiv:2106.00885*, 2021.

[169] Ilias Diakonikolas, Daniel M Kane, Alistair Stewart, and Yuxin Sun. Outlier-robust learning of ising models under dobrushin's condition. *arXiv preprint arXiv:2102.02171*, 2021.

[170] Ashish Katiyar, Jessica Hoffmann, and Constantine Caramanis. Robust estimation of tree structured gaussian graphical models. In *International Conference on Machine Learning*, pages 3292–3300. PMLR, 2019.

[171] Ashish Katiyar, Vatsal Shah, and Constantine Caramanis. Robust estimation of tree structured ising models. *arXiv preprint arXiv:2006.05601*, 2020.

[172] Ashish Katiyar, Soumya Basu, Vatsal Shah, and Constantine Caramanis. Recoverability landscape of tree structured markov random fields under symmetric noise. *arXiv preprint arXiv:2102.08554*, 2021.

[173] Marta Casanellas, Marina Garrote-López, and Piotr Zwiernik. Robust estimation of tree structured models. *arXiv preprint arXiv:2102.05472*, 2021.

[174] Steffen L. Lauritzen. *Graphical models*. Clarendon Press, Oxford :, 1996.

[175] Subhrajit Sinha, Pranav Sharma, Umesh Vaidya, and Venkataramana Ajjarapu. Identifying causal interaction in power system: Information-based approach. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017.

[176] Subhrajit Sinha, Pranav Sharma, Umesh Vaidya, and Venkataramana Ajjarapu. On information transfer based characterization of power system stability. *IEEE Transactions on Power Systems*, 2019.

[177] S Sinha and U Vaidya. On data-driven computation of information transfer for causal inference in discrete-time dynamical systems. *Journal of Nonlinear Science*, 2020.

[178] Jakob Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7):075310, 2018.

[179] V. R. Subramanian, A. Lamperski, and M. V. Salapaka. Network topology identification from corrupt data streams. In *IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1695–1700, 2017.

[180] V. R. Subramanian, A. Lamperski, and M. V. Salapaka. Inferring directed graphs for networks from corrupt data streams (in progress). In *IEEE 57th Annual Conference on Decision and Control (CDC)*, 2018.

[181] Frans MJ Willems, Yuri M Shtarkov, and Tjalling J Tjalkens. The context-tree weighting method: basic properties. *IEEE Transactions on Information Theory*, 41(3):653–664, 1995.

[182] Jiantao Jiao, Haim H. Permuter, Lei Zhao, Young-Han Kim, and Tsachy Weissman. Universal estimation of directed information. *IEEE Transactions on Information Theory*, 59(10):6220–6242, 2013.

[183] Gerhard Kramer. *Directed information for channels with feedback*. Hartung-Gorre, 1998.

[184] Trevor Werho, Vijay Vittal, Sharma Kolluri, and Sze Mei Wong. Power system connectivity monitoring using a graph theory network flow algorithm. *IEEE Transactions on Power Systems*, 31(6):4945–4952, 2016.

[185] Felipe Nunez, Yongqiang Wang, and Francis J Doyle. Synchronization of pulse-coupled oscillators on (strongly) connected graphs. *IEEE Transactions on Automatic Control*, 60(6):1710–1715, 2015.

[186] Sourav Patel, Sandeep Attree, Saurav Talukdar, Mangal Prakash, and Murti V Salapaka. Distributed apportioning in a power network for providing demand response services. In *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 38–44. IEEE, 2017.

[187] Sidhant Misra, Michael W Fisher, Scott Backhaus, Russell Bent, Michael Chertkov, and Feng Pan. Optimal compression in natural gas networks: A geometric programming approach. *IEEE transactions on control of network systems*, 2(1):47–56, 2014.

[188] L Sela Perelman and S Amin. Control of tree water networks: A geometric programming approach. *Water Resources Research*, 51(10):8409–8430, 2015.

[189] S. Talukdar, D. Deka, D. Materassi, and M. Salapaka. Exact topology reconstruction of radial dynamical systems with applications to distribution system of the power grid. In *2017 American Control Conference (ACC)*, pages 813–818, 2017.

[190] Lionel Barnett, Adam B Barrett, and Anil K Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Physical review letters*, 103(23):238701, 2009.

[191] V. R. Subramanian, A. Lamperski, and M. V. Salapaka. Corruption detection in networks of bi-directional dynamical systems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 4545–4550, 2019.

[192] Julius Adebayo, Taylor Southwick, Vasu Chetty, Enoch Yeung, Ye Yuan, Jorge Goncalves, Julianne Grose, John Prince, Guy-Bart Stan, and Sean Warnick. Dynamical structure function identifiability conditions enabling signal structure reconstruction. In *IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012.

[193] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

Appendix A

Network Structure Identification in LTI Systems using Corrupt Data Streams

A.1 Proof of Theorem 2

Define the following deviations from the mean: $\Delta A_i[t] = A_i[t] - \bar{A}_i$, $\Delta B_i[t] = B_i[t] - \bar{B}_i$, $\Delta C_i[t] = C_i[t] - \bar{C}_i$, and $\Delta D_i[t] = D_i[t] - \bar{D}_i$

Note that the Lyapunov equation, (2.9), can be expressed as:

$$P = \bar{A}_i^\top P \bar{A}_i + \mathbb{E}[\Delta A_i[t]^\top P \Delta A_i[t]] + Q \succeq \bar{A}_i^\top P \bar{A}_i + Q. \tag{A.1}$$

Here $S \preceq T$ denotes that $T - S$ is positive semidefinite. Since a solution must hold for all Q, it must hold, in particular for positive definite Q. Thus, $\bar{A}_i$ must be a stable matrix.

Set $\bar{u}_i[t] = (h_i \star y_i)[t] = \mathbb{E}[u_i[t]|y_i]$, so that $\Delta u_i[t] = u_i[t] - \bar{u}_i[t]$.

With this notation, the cross spectrum, (2.10b), will be derived:

$$\begin{aligned} R_{u_i y_i}[t] &= \mathbb{E}[u_i[t] y_i[0]] && \text{(A.2)} \\ &= \mathbb{E}[\mathbb{E}[u_i[t] y_i[0] | y_i]] && \text{(A.3)} \\ &= \mathbb{E}[(h_i \star y_i)[t] y_i[0]] && \text{(A.4)} \\ &= (h_i \star R_{y_i y_i})[t]. && \text{(A.5)} \end{aligned}$$

Here, (A.3) is due to the tower property of conditional expectation. Then (2.10b) follows by taking Z-transforms.

Since $\bar{A}_i$ is stable and $y_i[t]$ is wide-sense stationary, we must have that $\bar{u}_i[t]$ is wide-sense stationary.

Note that by construction, $R_{u_iu_i}[t] = R_{\bar{u}_i\bar{u}_i}[t] + R_{\Delta u_i \Delta u_i}[t]$. Furthermore, we must have that

$$R_{\bar{u}_i\bar{u}_i}[t] = (h_i \star R_{yy} \star h_i^*)[t], \tag{A.6}$$

where h_i^* is the time-reversed, transposed impulse response. Thus, (2.10a) holds by taking Z-transforms.

The only part that remains to be proved is that u_i is wide-sense stationary. This will follow as long as long as $\Delta u_i[t]$ has a finite autocorrelation.

To show that $R_{\Delta u_i \Delta u_i}[t]$ is bounded, we will explicitly construct an expression for it. To derive this expression, we need expressions for the autocorrelation of x_i and the cross correlation between x_i and y_i.

Let $\bar{x}_i[t] = \left(\left[\begin{array}{c|c} \bar{A}_i & \bar{B}_i \\ \hline I & 0 \end{array}\right] \star y_i\right)[t]$ and let $\Delta x_i[t] = x_i[t] - \bar{x}_i[t]$. Note that $\bar{x}_i[t] = \mathbb{E}[x_i[t]|y_i]$. As with $\bar{u}_i$, we have that $\bar{x}_i[t]$ is wide-sense stationary. Using a derivation identical to that of $R_{u_iy_i}[t]$, we have that the cross correlation of x_i and y_i is given by:

$$R_{x_iy_i}[t] = \left(\left[\begin{array}{c|c} \bar{A}_i & \bar{B}_i \\ \hline I & 0 \end{array}\right] \star R_{y_iy_i}\right)[t] \tag{A.7}$$

Thus, we see that $R_{x_iy_i}[t] = R_{\bar{x}_iy_i}[t]$.

Now we will work out the autocorrelation of x_i. The autocorrelation of $\bar{x}_i[t]$ is given by:

$$R_{\bar{x}_i\bar{x}_i}[t] = \left(\left[\begin{array}{c|c} \bar{A}_i & \bar{B}_i \\ \hline I & 0 \end{array}\right] \star R_{y_iy_i} \star \left[\begin{array}{c|c} \bar{A}_i & \bar{B}_i \\ \hline I & 0 \end{array}\right]^*\right)[t]. \tag{A.8}$$

By construction, we have that $R_{x_ix_i}[t] = R_{\bar{x}_i\bar{x}_i}[t] + R_{\Delta x_i \Delta x_i}[t]$. The following lemma characterizes the autocorrelations of $\Delta x_i[k]$.

Lemma 6. *Assume that a solution to the generalized Lyapunov equation,* (2.9)*, holds*

for all Q. Then $R_{\Delta x_i \Delta x_i}[0]$ is uniquely defined by:

$$R_{\Delta x_i \Delta x_i}[0] = \mathbb{E}[A_i[0] R_{\Delta x_i \Delta x_i}[0] A_i[0]^\top] + W + \mathbb{E}\left[\begin{bmatrix} \Delta A_i[0] & \Delta B_i[0] \end{bmatrix} \begin{bmatrix} R_{\bar{x}_i \bar{x}_i}[0] & R_{\bar{x}_i y_i}[0] \\ R_{y_i \bar{x}_i}[0] & R_{y_i y_i}[0] \end{bmatrix} \begin{bmatrix} \Delta A_i[0]^\top \\ \Delta B_i[0]^\top \end{bmatrix} \right]. \quad \text{(A.9)}$$

For $k > 0$,

$$R_{\Delta x_i \Delta x_i}[k] = \bar{A}_i^k R_{\Delta x_i \Delta x_i}[0]$$
$$R_{\Delta x_i \Delta x_i}[-k] = R_{\Delta x_i \Delta x_i}[k]^\top.$$

Proof. For $k > 0$ we have

$$\begin{aligned} &R_{\bar{x}_i \bar{x}_i}[k] + R_{\Delta x_i \Delta x_i}[k] \\ &\quad = \mathbb{E}[x_i[k] x_i[0]^\top] \\ &\quad = \mathbb{E}[(A_i[k-1] x_i[k-1] + B_i[k-1] y_i[k-1]) x_i[0]^\top] \\ &\quad = \bar{A}_i R_{x_i x_i}(k-1) + \bar{B}_i R_{y_i x_i}(k-1) \\ &\quad = (\bar{A}_i R_{\bar{x}_i \bar{x}_i}(k-1) + \bar{B}_i R_{y_i x_i}(k-1)) \\ &\quad + \bar{A}_i R_{\Delta x_i \Delta x_i}(k-1) \\ &\quad = R_{\bar{x}_i \bar{x}_i}[k] + \bar{A}_i R_{\Delta x_i \Delta x_i}[k-1]. \end{aligned}$$

Thus, the formula for $R_{\Delta x_i \Delta x_i}[k]$ holds for $k \neq 0$. (The expression for $k < 0$ follows from transposing.)

Note that

$$\begin{aligned} &\Delta x_i[k+1] \\ &\quad = (\bar{A}_i + \Delta A_i[k])(\bar{x}_i[k] + \Delta x_i[k]) + (\bar{B}_i + \Delta B_i[k]) y_i[k] \\ &\quad + w_i[k] - \bar{A}_i \bar{x}_i[k] - \bar{B}_i y_i[k] \\ &\quad = A_i[k] \Delta x_i[k] + \Delta A_i[k] \bar{x}_i[k] + \Delta B_i[k] y_i[k] + w_i[k]. \end{aligned}$$

Furthermore, note that $\Delta x_i[k]$ is independent of $\Delta A_i[k]$ and $\Delta B_i[k]$. The expression for $R_{\Delta x_i \Delta x_i}(0)$ follows by setting $\mathbb{E}[\Delta x_i[k+1] \Delta x_i[k+1]^\top] = \mathbb{E}[\Delta x_i[k] \Delta x_i[k]^\top]$.

Note that $R_{\Delta x_i \Delta x_i}(0)$ can be computed from (2.9) with

$$Q = W + \mathbb{E}\left[\begin{bmatrix}\Delta A_i[0] & \Delta B_i[0]\end{bmatrix}\begin{bmatrix}R_{\bar{x}_i\bar{x}_i}(0) & R_{\bar{x}_i y_i}(0)\\ R_{y_i\bar{x}_i}(0) & R_{y_i y_i}(0)\end{bmatrix}\begin{bmatrix}\Delta A_i[0]^\top\\ \Delta B_i[0]^\top\end{bmatrix}\right] \tag{A.10}$$

□

As discussed above, the proof of the theorem will be completed once the autocorrelation of Δu_i is characterized. The following lemma gives the desired characterization.

Lemma 7. *For $k = 0$, $R_{\Delta u_i \Delta u_i}[0]$ is given by*

$$R_{\Delta u_i \Delta u_i}[0] = \bar{C}_i R_{\Delta x_i \Delta x_i}[0]\bar{C}_i^\top + V + \mathbb{E}\left[\begin{bmatrix}\Delta C_i[0] & \Delta D_i[0]\end{bmatrix}\begin{bmatrix}R_{x_i x_i}[0] & R_{x_i y_i}[0]\\ R_{y_i x_i}[0] & R_{y_i y_i}[0]\end{bmatrix}\begin{bmatrix}\Delta C_i[0]^\top\\ \Delta D_i[0]^\top\end{bmatrix}\right] \tag{A.11}$$

For $k > 0$, $R_{\Delta u_i \Delta u_i}[k]$ is given by

$$R_{\Delta u_i \Delta u_i}[k] = \bar{C}_i R_{\Delta x_i \Delta x_i}[k]\bar{C}_i^\top + \bar{C}_i \bar{A}_i^{k-1} S + \bar{C}_i \bar{A}_i^{k-1}\mathbb{E}\left[\begin{bmatrix}\Delta A_i[0] & \Delta B_i[0]\end{bmatrix}\begin{bmatrix}R_{x_i x_i}[0] & R_{x_i y_i}[0]\\ R_{y_i x_i}[0] & R_{y_i y_i}[0]\end{bmatrix}\cdot \begin{bmatrix}\Delta C_i[0]^\top\\ \Delta D_i[0]^\top\end{bmatrix}\right] \tag{A.12}$$

For $k < 0$, $R_{\Delta u_i \Delta u_i}[k] = R_{\Delta u_i \Delta u_i}[-k]$.

Proof. Note that $\Delta u_i[k]$ can be decomposed as:

$$\begin{aligned}
&\Delta u_i[k] && \text{(A.13)}\\
&= u_i[k] - \bar{u}_i[k] && \text{(A.14)}\\
&= (\bar{C}_i + \Delta C_i[k])(\bar{x}_i[k] + \Delta x_i[k]) + (\bar{D}_i + \Delta D_i[k])y_i[k] && \text{(A.15)}\\
&+ v_i[k] - \bar{C}_i\bar{x}_i[k] - \bar{D}_i y_i[k] && \text{(A.16)}\\
&= \bar{C}_i\Delta x_i[k] + \Delta C_i[k]x_i[k] + \Delta D_i[k]y_i[k] + v_i[k] && \text{(A.17)}
\end{aligned}$$

As before, $\Delta x_i[k]$ is independent of $\Delta C_i[k]$ and $\Delta D_i[k]$. Thus, the expression for $R_{\Delta u_i \Delta u_i}[0]$ follows by computing $\mathbb{E}[\Delta u_i[k]^2]$.

For $k > 0$, note that $\Delta C_i[k]$ and $\Delta D_i[k]$ are independent of $\Delta C_i[0]$ and $\Delta D_i[0]$. However, $\Delta x_i[k]$ may be correlated with $\Delta C_i[0]$, $\Delta D_i[0]$, and $v_i[0]$. So, multiplying the expression from (A.17) for $k > 0$ and $k = 0$ and dropping the $\Delta C_i[k]$ and $\Delta D_i[k]$ terms gives

$$\begin{aligned} R_{\Delta u_i \Delta u_i}(k) &= \mathbb{E}\left[\bar{C}_i \Delta x_i[k](\bar{C}_i \Delta x_i[0]+)^\top\right] \\ &\quad + \mathbb{E}\left[\bar{C}_i \Delta x_i[k]\left(\Delta C_i[0] x_i[0] + \Delta D_i[0] y_i[0] + v_i[0]\right)^\top\right] && \text{(A.18)} \\ &= \bar{C}_i R_{\Delta x_i \Delta x_i}(k) \bar{C}_i^\top && \text{(A.19)} \\ &\quad + \bar{C}_i \mathbb{E}[\Delta x_i[k](\Delta C_i[0] x_i[0] + \Delta D_i[0] y_i[0] + v_i[0])^\top] && \text{(A.20)} \end{aligned}$$

Let $A_i[j:k]$ be the product defined by $A_i[k:k] = I$ and $A_i[j:k] = A_{i,j}[k-1]A_i[k-2]\cdots A_i[j]$ for $j < k$. An induction argument shows that

$$\begin{aligned} x_i[k] &= A_i[0:k]x_i[0] + \sum_{j=0}^{k-1} A_i[j+1:k](B_i[j]y_i[j] + w_i[j]) \\ &= A_i[1:k]A_i[0]x_i[0] + B_i[0]y_i[0] + w_i[0]) + \\ &\quad + \sum_{j=1}^{k-1} A_i[j+1:k](B_i[j]y_i[j] + w_i[j]). \end{aligned}$$

Let $\mathcal{F}$ be the σ-algebra generated by y_i and all of the random terms: $A_i[j], B_i[j], C_i[j], D_i[j], w_i[j], v_i[j]$ for $i \leq 0$. Then the expression for $x_i[k]$ implies that

$$\begin{aligned} \mathbb{E}[x_i[k]|\mathcal{F}] &= \sum_{j=1}^{k-1} \bar{A}^{k-1-j}\bar{B}y_i[j] \\ &\quad + \bar{A}^{k-1}\left((\bar{A} + \Delta A_i[0])x_i[0] + (\bar{B} + \Delta B_0)y_i[0] + w_i[0]\right) \\ &= \bar{x}_i[k] + \bar{A}^{k-1}\bar{A}\Delta x_i[0] + \\ &\quad + \bar{A}^{k-1}(\Delta A_i[0]x_i[0] + \Delta B_i[0]y_i[0] + w_i[0]). \end{aligned}$$

Using the tower property gives:

$$\begin{aligned}
&\mathbb{E}[\Delta x_i[k](\Delta C_i[0]x_i[0] + \Delta D_i[0]y_i[0] + v_i[0])^\top] \\
&\quad = \mathbb{E}[\mathbb{E}[\Delta x_i[k](\Delta C_i[0]x_i[0] + \Delta D_i[0]y_i[0] + v_i[0])^\top | \mathcal{F}]] \\
&\quad = \bar{A}^{k-1}\mathbb{E}[(\Delta A_i[0]x_i[0] + \Delta B_i[0]y_i[0] + w_i[0]) \\
&\qquad \cdot (\Delta C_i[0]x_i[0] + \Delta D_i[0]y_i[0] + v_i[0])^\top],
\end{aligned}$$

where the last equality used that $\Delta x_i[0]$ is independent of $\Delta A_i[0]$, $\Delta B_i[0]$, and $v_i[0]$. Combining this result with (A.19) gives the desired expression for $R_{\Delta u \Delta u}(k)$. The expression for $R_{\Delta u \Delta u}(-k)$ follows because $\Delta u_i[k]$ is a real scalar. □

A.2 Proof of Theorem 3

First, we will describe the structure of $\Phi_{uu}(z)$. For compact notation, we will often drop the z arguments.

For $p = 1, \ldots, N$, if p is not a perturbed node, set $H_p(z) = 1$ and $\theta_p(z) = 0$. With this notation, (2.10) implies that the entries of Φ_{uu} are given by:

$$(\Phi_{uu})_{pq} = \begin{cases} H_p(\Phi_{yy})_{pq}H_q^* & \text{if } p \neq q \\ H_p(\Phi_{yy})_{pp}H_p^* + \theta_p & \text{if } p = q \end{cases}$$

When $p \neq q$, there is no θ term because the perturbations were assumed to be independent.

In matrix notation, we have that:

$$\Phi_{uu} = H\Phi_{yy}H^* + \sum_{k=1}^{n} \mathcal{D}_{v_k}$$

where H is the diagonal matrix with entries H_p on the diagonal and $\mathcal{D}_{v_k}(z) = b_{v_k}\theta_{v_k}(z)b_{v_k}^T$ where b_{v_k} is the canonical unit vector with 1 at entry v_k.

Set $\Psi_0 = H\Phi_{yy}H^*$. For $k = 0, \ldots, n-1$, we can inductively define the matrices:

$$\Psi_{k+1} = \Psi_k + b_{v_{k+1}}\theta_{v_{k+1}}b_{v_{k+1}}^T \tag{A.21}$$

For $k = 1, \ldots, n$ let $Z_k = \{v_1, \ldots, v_k\}$ and let $G_{Z_k}^M$ be the perturbed graph constructed by adding edges $i - j$ to the original moral graph if there is a path from i to j whose intermediate nodes are all in Z_k.

We will inductively prove the following claim: For $k = 1, \ldots, n$, if $(\Psi_k^{-1})_{pq} \neq 0$, then p and q are neighbors in $G^M_{Z_k}$. Proving this claim is sufficient to prove the theorem, since $\Psi_n = \Phi_{uu}$ and $Z_n = Z$.

First we focus on the $k = 1$ case. Using the Woodbury Matrix identity we have, $\Psi_1^{-1} = \Psi_0^{-1} - \Gamma_1$, where $\Gamma_1 := (\Psi_0^{-1} b_{v_1} b_{v_1}^T \Psi_0^{-1})\Delta_{v_1}^{-1}$ and $\Delta_{v_1} = \theta_{v_1}^{-1} + b_{v_1}^T \Psi_0^{-1}(z) b_{v_1}$ is a scalar. Therefore, $(\Psi_1^{-1})_{pq} = (\Psi_0^{-1})_{pq} - (\Gamma_1)_{pq}$.

If $(\Psi_1^{-1})_{pq} \neq 0$ then at least one of the conditions (i) $(\Psi_0^{-1})_{pq} \neq 0$ or (ii) $(\Gamma_1)_{pq} \neq 0$ must hold.

Suppose that $(\Psi_0^{-1})_{pq} \neq 0$. Then $(H^{-*}(z)\Phi_{yy}^{-1} H^{-1}(z))_{pq} \neq 0$. As H is diagonal it follows that $(\Phi_{yy}^{-1})_{pq} \neq 0$. From Corollary 1, it follows that p and q are neighbors in G^M. Thus p and q are neighbors in $G^M_{B_1}$.

Suppose that $(\Gamma_1)_{pq} \neq 0$. Then it follows that $(\Psi_0^{-1} b_{v_1} b_{v_1}^T \Psi_0^{-1})_{pq} \Delta_{v_1}^{-1} \neq 0$. Thus $(\Psi_0^{-1} b_{v_1})_p \neq 0$ and $(b_{v_1}^T \Psi_0^{-1})_q \neq 0$. Noting that $\Psi_0 = H\Phi_{yy}H^*$, it follows that , $(\Phi_{yy}^{-1})_{pv_1} \neq 0$ and $(\Phi_{yy}^{-1})_{v_1 q} \neq 0$. From Corollary 1 it follows that $v_1 - p$ and $v_1 - q$ are edges in the moral graph G^M. Thus, there is a path from p to q whose only intermediate node is $v_1 \in Z_1$. Thus, p, q are neighbors in $G^M_{Z_1}$ and the claim is verified for $k = 1$.

Now assume that the claim holds for some $k > 1$. Combining the Woodbury matrix identity with (F.1) implies that

$$\Psi_{k+1}^{-1} = \Psi_k^{-1} - \Gamma_{k+1}$$

where $\Gamma_{k+1} = \Psi_k^{-1} b_{v_{k+1}} b_{v_{k+1}}^T \Psi_k^{-1} \Delta_{v_{k+1}}^{-1}$ and $\Delta_{v_{k+1}} = \theta_{v_{k+1}}^{-1} + b_{v_{k+1}}^T \Psi_k^{-1}(z) b_{v_{k+1}}$ is a scalar.

As before, if $(\Psi_{k+1}^{-1})_{pq} \neq 0$, then either $(\Psi_k^{-1})_{pq} \neq 0$ or $(\Gamma_{k+1})_{pq} \neq 0$.

If $(\Psi_k^{-1})_{pq} \neq 0$, then the induction hypothesis implies that p and q are neighbors in $G^M_{Z_k}$. Since $Z_k \subset Z_{k+1}$, it follows that p and q are neighbors in $G^M_{Z_{k+1}}$.

If $(\Gamma_{k+1})_{pq} \neq 0$, then as in the $k = 1$ case, we must have that $(\Psi_k^{-1})_{pv_{k+1}} \neq 0$ and $(\Psi_k^{-1})_{v_{k+1}q} \neq 0$. This implies that $p - v_{k+1} \in A^M_{Z_k}$ and $v_{k+1} - q \in A^M_{Z_k}$. Thus, either p and v_{k+1} are kins in the original moral graph, or there is a path from p to v_{k+1} whose intermediate nodes are in Z_k. Similarly, for q and v_{k+1}. It follows that there is a path from p to q whose nodes are in Z_{k+1}, and thus p and q are neighbors in $G^M_{Z_{k+1}}$. The claim, and thus the theorem, are now proved. □

Appendix B

Spurious Correlations in Perturbed Markov Random Fields

B.1 Proof for Lemma 1

We will prove the lemma for discrete random variables. The proof for continuous random variables is identical except that marginalization would be represented by integrals instead of sums.

Let $Z = \{v_1, \ldots, v_n\}$, $Z_0 = \emptyset$ and $Z_k = \{v_1, \ldots, v_k\}$. We will prove inductively that $(Y_{\bar{Z}_k}, U_{Z_k})$ factorizes with respect to G_{Z_k}.

The base case with $Z_0 = \emptyset$ is immediate since $(Y_{\bar{Z}_0}, U_{Z_0}) = Y$ and $G_{Z_0} = G$. Now assume inductively that $(Y_{\bar{Z}_{k-1}}, U_{Z_{k-1}})$ factorizes with respect to $G_{Z_{k-1}}$ for some $k \geq 1$.

$$P(Y_{\bar{Z}_k}, U_{Z_k}) = \sum_{Y_{Z_k}} \prod_{q \in Q} \Psi_q(Y_q) \prod_{i=1}^{k} \Psi_{v_i,(v_i)_p}(y_{v_i}, u_{v_i}) \tag{B.1}$$

$$= \sum_{y_{v_k}} \left(\sum_{Y_{Z_{k-1}}} \prod_{q \in Q} \Psi_q(Y_q) \prod_{i=1}^{k-1} \Psi_{v_i,(v_i)_p}(y_{v_i}, u_{v_i}) \right) \cdot \Psi_{v_k,(v_k)_p}(y_{v_k}, u_{v_k})$$

$$= \sum_{y_{v_k}} P(Y_{\bar{Z}_{k-1}}, U_{Z_{k-1}}) \Psi_{v_k,(v_k)_p}(y_{v_k}, u_{v_k}) \tag{B.2}$$

The first line, (B.1) follows by marginalizing Y_{Z_k} out of the factorized distribution from (3.4). Then the terms are regrouped and then (B.1) is employed for $P(Y_{\bar{Z}_{k-1}}, U_{Z_{k-1}})$.

By induction, we have that $P(Y_{\bar{Z}_{k-1}}, U_{Z_{k-1}})$ factorizes according to a collection of cliques, C, in G_{Z_k}. Let $C_{v_k} \subset C$ be the collection of cliques such that $v_k \in c$ for all $c \in C_{v_k}$. For compact notation, let $X = (Y, U_Z)$. Then the formula for $P(Y_{\bar{Z}_k}, U_{Z_k})$ can be expressed as

$$P(Y_{\bar{Z}_k}, U_{Z_k}) \quad = \quad \left(\sum_{y_{v_k}} \prod_{c \in C_{v_k}} \Psi_c(X_c) \Phi_{v_k,(v_k)_p}(y_{v_k}, u_{v_k}) \right) \cdot \prod_{c \in C \setminus C_{v_k}} \Phi_c(X_c). \tag{B.3}$$

The second term on the right is a collection of factors corresponding to cliques in $G_{Z_{k-1}}$. Now, since $A_{Z_{k-1}} \subset A_{Z_k}$, they must also be cliques of G_{Z_k}. The lemma will be proved if the variables first term on the right correspond to a clique in G_{Z_k}.

Say $i \neq v_k$ and $j \neq v_k$ are nodes corresponding to variables in the sum in (B.3). Then there must be paths from i to v_k and v_k to j such that any intermediate node is in Z_{k-1}. Now, since $v_k \in Z_k$, there is a path from i to j such that all of the intermediate nodes are in Z_k. Thus, the nodes in the sum form a clique in G_{Z_k}.

B.2 Proof for Theorem 4

From lemma 1 we know that U factorizes according to G_Z. Thus, positivity of $P(U)$ implies that the pairwise Markov property is equivalent to U factorizing according to G_Z. See [174]. Therefore, $\text{sep}(i, j | V \setminus \{i, j\})$ (in G_Z) if and only if $u_i \perp\!\!\!\perp u_j \mid U_{\bar{\bar{ij}}}$. Note that $\text{sep}(i, j | V \setminus \{i, j\})$ means precisely that $i - j \notin A_Z$.

Appendix C

Effects of Data Corruption on Network Identification for Nonlinear Systems

C.1 Proof for Theorem 6

We will show that if $i \to j \notin A_Z$, then there is no trail between $W_i^{(t-1)}$ and $W_j[t]$ that is active given $\{W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\}$ in the PDBN G'_Z, for all $t > 0$. However, trail in G'_Z between $W_i^{(t-1)}$ and $W_j[t]$ can contain future state node $\alpha_{b_m}[t_m]$ such that $t_m \geq t$ and $b_m \in V$. Note that such nodes are not observed in $\{W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\}$ and can therefore make the trail in PDBN active. However, the following lemma proves that such nodes actually makes the trail inactive given $\{W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\}$.

Lemma 8. *Consider a generative graph, $G = (V, A)$, consisting of N nodes. Let $Z = \{v_1, \ldots, v_n\} \subset V$ be the set of n perturbed nodes where each perturbation is described by (4.4). Denote the data-streams as follows: $U_Z := \{U_i\}_{i \in Z}$ and $Y_{\bar{Z}} := \{Y_j\}_{j \in \bar{Z}}$ where $\bar{Z} = V \setminus Z$. Let the measured data-streams be $\mathcal{W} = U_Z \cup Y_{\bar{Z}} = \{W_1, W_2, \ldots, W_N\}$. Let $G' = (V', A')$ be the dynamic Bayesian network (DBN) associated with G and $G'_Z = (V'_Z, A'_Z)$ be the perturbed DBN. If $i \to j \notin A$ and if a trail in G'_Z between $W_i^{(t-1)}$ and $W_j[t]$ contains a node $\alpha_{b_m}[t_m]$ such that $t_m \geq t$ and $b_m \in V$, then for all $t > 0$, the trail is not active given $\{W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\}$.*

Proof. Consider any trail from a node in $W_i^{(t-1)}$ to $W_j[t]$ in G'_Z. Denote this by $trlG'_Z := W_i[t_1] = \alpha_{b_1}[t_1] - \alpha_{b_2}[t_2] - \cdots - \alpha_{b_{r-1}}[t_{r-1}] - \alpha_{b_r}[t_r] = W_j[t]$ where $0 \leq t_1 < t$. Here, b_k denotes the corresponding vertex in V for $k = \{1, 2, \ldots, r\}$. Also, $\alpha_{b_k}[t_k] = U_{b_k}[t_k]$ if $b_k \in Z$ or $\alpha_{b_k}[t_k] = Y_{b_k}[t_k]$ otherwise. For compact notation, set $\theta := \{W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\}$.

The trail has length at least 3. As $i \to j \notin A$ and if $j \notin Z$, then $Y_j[t]$ does not dynamically depend on process Y_i and clearly not on U_i. If $j \in Z$, then by (4.4), $U_j[t]$ does not dynamically depend on Y_i nor U_i. Thus, there is no direct link of the form $\alpha_i[t'] \to \alpha_j[t"]$ in G'_Z, for any $t', t"$. In particular, $W_i[t_1] \to W_j[t] \notin G'_Z$. Thus, there are at least 3 nodes in the trail, $trlG'_Z$.

Unobserved collider in trail. Without loss of generality, choose $t_m = \max\{t_1, \ldots, t_{r-1}\} \geq t$. Consider the sub-trail $subtrl' := \alpha_{b_{m-1}}[t_{m-1}] - \alpha_{b_m}[t_m] - \alpha_{b_{m+1}}[t_{m+1}]$ of $trlG'_Z$. By maximality of t_m, $t_m \geq t_{m-1}$ and $t_m \geq t_{m+1}$. We will show that one of $\alpha_{b_{m-1}}[t_{m-1}]$, $\alpha_{b_m}[t_m]$, and $\alpha_{b_{m+1}}[t_{m+1}]$ is a collider not in θ and therefore the trail $trlG'_Z$ cannot be active given θ.

Suppose $t_m > t_{m-1}$ and $t_m > t_{m+1}$. Then, $subtrl'$ is of the form, $\alpha_{b_{m-1}}[t_{m-1}] \to \alpha_{b_m}[t_m] \leftarrow \alpha_{b_{m+1}}[t_{m+1}]$. Note that, as $t_m \geq t$, it follows that neither $\alpha_{b_m}[t_m]$ nor any of its descendants can be in θ and hence not observed.

Now, consider $t_m > t_{m-1}$ and $t_m = t_{m+1}$. (The case of $t_m > t_{m+1}$ and $t_m = t_{m-1}$ can be proven similarly). By the generative model in (4.1), by strict causality, for any node $p \in V$, $Y_p[t_p]$ does not dynamically depend on any $Y_q[t_p]$ for $q \in \{p, \mathcal{P}'(p)\}$. By the perturbation model described by (4.4), for any $q \in Z$, $U_q[t_q]$ dynamically depends only on $\{U_q^{(t_q-1)}, Y_q^{(t_q)}\}$. As $t_m = t_{m+1}$, we therefore have $b_m = b_{m+1}$ such that $b_m \in Z$ and, one of $\alpha_{b_m}[t_m]$ and $\alpha_{b_{m+1}}[t_{m+1}]$ is actually a perturbed measurement $U_{b_m}[t_m]$ while the other being $Y_{b_m}[t_m]$.

Suppose $\alpha_{b_m}[t_m] = U_{b_m}[t_m]$. Then, $\alpha_{b_{m+1}}[t_{m+1}] = Y_{b_m}[t_m]$. As $t_m > t_{m-1}$, $subtrl'$ is in fact $\alpha_{b_{m-1}}[t_{m-1}] \to \alpha_{b_m}[t_m] = U_{b_m}[t_m] \leftarrow \alpha_{b_{m+1}}[t_{m+1}] = Y_{b_m}[t_m]$. Therefore, $\alpha_{b_m}[t_m]$ is a collider and as $t_m \geq t$, it is not observed in θ.

Suppose instead that $\alpha_{b_m}[t_m] = Y_{b_m}[t_m]$. Then, $\alpha_{b_{m+1}}[t_{m+1}] = U_{b_m}[t_m]$. As $b_m \in Z$ and maximality of t_m implies $\alpha_{b_{m+2}}[t_{m+2}] \in \{U_{b_m}^{(t_m-1)}, Y_{b_m}^{(t_m-1)}\}$. Thus, we have $\alpha_{b_{m-1}}[t_{m-1}] - \alpha_{b_m}[t_m] = Y_{b_m}[t_m] \to \alpha_{b_{m+1}}[t_{m+1}] = U_{b_m}[t_m] \leftarrow \alpha_{b_{m+2}}[t_{m+2}]$ in $trlG'_Z$. Therefore, $\alpha_{b_{m+1}}[t_{m+1}]$ is a collider not observed in θ. □

Proof of Theorem 6

For rest of the proof, denote $\theta := \{W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\}$. Note that if $i \to j \notin A_Z$, then there is no directed edge from i to j in G, and every trail from i to j in G violates at least one of the conditions of Definition 26. We will consider these cases separately and show that no active trail exists in G'_Z in each case. Denote any trail connecting a node in $W_i^{(t-1)}$ and $W_j[t]$ in G'_Z, by $trlG'_Z := W_i[t_1] = \alpha_{b_1}[t_1] - \alpha_{b_2}[t_2] - \cdots - \alpha_{b_{r-1}}[t_{r-1}] - \alpha_{b_r}[t_r] = W_j[t]$ where $0 \leq t_1 < t$ and b_k denotes the corresponding vertex in V for $k = \{1, 2, \ldots, r\}$. Here, $\alpha_v[t_v] = U_v[t_v]$ if $v \in Z$ or $\alpha_v[t_v] = Y_v[t_v]$ otherwise. Using Lemma 8, if any t' in $\{t_2, \ldots, t_{r-1}\}$ is such that $t' \geq t$, then $trlG'_Z$ is not active. Now, consider $0 \leq t_1, t_2, t_3, \ldots, t_{r-1} < t$. We will first show that any such trail in G'_Z, $trlG'_Z$, can be mapped to a trail in G, $trlG := i = v_1 - v_2 - v_3 \ldots v_{k-1} - v_k = j$ as follows:

Initialize: $k = 1$ and $v_1 = b_1$.
for $l = 1 : r - 1$ **do**
  if $b_{l+1} \neq b_l$ in $\alpha_{b_l}[t_l] - \alpha_{b_{l+1}}[t_{l+1}]$ along $trlG'_Z$ **then**
    Set $v_{k+1} = b_{l+1}$.
    Add edge $v_k - v_{k+1}$ with the same direction as
    $\alpha_{b_l}[t_l] - \alpha_{b_{l+1}}[t_{l+1}]$.
    Set $s_k = t_l$ and $\tau_{k+1} = t_{l+1}$
    Set $k = k + 1$
  end if
end for

Additionally, note that $v_k - v_{k+1}$ corresponds to an edge $\alpha_{v_k}[s_k] - \alpha_{v_{k+1}}[\tau_{k+1}]$ in G'_Z.

Now, let us reason out why such a construction is always feasible. To this, we claim that for any successive pair $\alpha_{b_l}[t_l] - \alpha_{b_{l+1}}[t_{l+1}]$, either $b_l = b_{l+1}$ or, $b_l \neq b_{l+1}$ and $b_l - b_{l+1} \in A$ with the same direction as in $\alpha_{b_l}[t_l] - \alpha_{b_{l+1}}[t_{l+1}]$. Assume $\alpha_{b_l}[t_l] \to \alpha_{b_{l+1}}[t_{l+1}]$. (The case of $\alpha_{b_l}[t_l] \leftarrow \alpha_{b_{l+1}}[t_{l+1}]$ is similar). Then, either $t_l = t_{l+1}$ or $t_l < t_{l+1}$. Consider, $t_l = t_{l+1}$. Then, the link must have the form $Y_{b_l}[t_l] \to U_{b_l}[t_l]$, as this is the only instantaneous influence defined in (4.1) or (4.4). Thus, $b_l = b_{l+1}$ in this case.

Suppose, $t_l < t_{l+1}$. Either, $b_{l+1} \in Z$ or $b_{l+1} \notin Z$. Consider $b_{l+1} \in Z$. By the

perturbation model described by (4.4), $\alpha_{b_l}[t_l] \in \{Y_{b_{l+1}}^{(t_{l+1}-1)}, U_{b_{l+1}}^{(t_{l+1}-1)}\}$. Therefore, $b_l = b_{l+1}$. Suppose, $b_{l+1} \notin Z$. Then, $\alpha_{b_{l+1}}[t_{l+1}] = Y_{b_{l+1}}[t_{l+1}]$. By the generative model in (4.1), we either have dynamic dependence on self-history or history of other nodes. That is, $\alpha_{b_l}[t_l] \in \{Y_{b_l}^{(t_{l+1}-1)}, \underset{q\in\mathcal{P}'(b_{l+1})}{\cup} Y_q^{(t_{l+1}-1)}\}$. Then, $b_l = b_{l+1}$ when there is dependence on self-history. Otherwise, $b_l \in \mathcal{P}'(b_{l+1})$. Thus, $b_l \to b_{l+1} \in A$. Let us consider an example- from a trail of the form $U_1[t_1] \leftarrow Y_1[t_2] \leftarrow Y_2[t_3] \to Y_3[t_4] \to Y_3[t_5] \to U_3[t]$ in G'_Z, a trail $trlG$ in G can be constructed as $1 \leftarrow 2 \to 3$.

Additionally, we may assume that for $m = 2, \cdots, r-1$ we have that $\alpha_{b_m}[t_m] \neq W_i[t_m]$ in $trlG'_Z$. If $\alpha_{b_m}[t_m] = W_i[t_m]$ for some $m > 1$, then the sub-trail of $trlG'_Z$, $W_i[t_m] = \alpha_{b_m}[t_m] - \alpha_{b_{m+1}}[t_{m+1}] - \cdots - \alpha_r[t_r] = W_j[t]$ is a trail from $W_i[t_m] \in W_i^{(t-1)}$ to $W_j[t]$. This trail is of strictly shorter length than $trlG'_Z$. Thus, if the shorter trail cannot be active then the longer trail, $trlG'_Z$, cannot be active either. Also, by following the construction procedure described above, this condition implies that $v_l \neq i$ for $l = 2, 3, \cdots, k$ in $trlG$. Call this condition $loop_i$. To summarize, let $trlG := i = v_1 - v_2 - v_3 \ldots v_{k-1} - v_k = j$ be any trail connecting i and j in G constructed by following the above procedure from the trail $trlG'_Z$: $W_i[t_1] = \alpha_{b_1}[t_1] - \alpha_{b_2}[t_2] - \cdots - \alpha_{b_{r-1}}[t_{r-1}] - \alpha_{b_r}[t_r] = W_j[t]$. Since, $i \to j \notin A_Z$, this trail must violate any of the conditions P1), P2) and P3). We will now consider these cases separately and prove that there is no corresponding active trail in G'_Z.

If condition P1) is violated, then $trlG$ must have that $j \notin Z$ and $v_{k-1} \leftarrow j$. In this case, $W_j = Y_j$. Then, either $b_{r-1} = j$ or $b_{r-1} \neq j$. By construction of $trlG$, if $b_{r-1} \neq j$, then $b_{r-1} = v_{k-1}$. As $v_{k-1} \leftarrow j$, we must then have $\alpha_{b_{r-1}}[t_{r-1}] \leftarrow \alpha_{b_r}[t_r]$. However, this implies $t_r = t < t_{r-1}$ which violates the condition that $0 \leq t_1, t_2, t_3, \ldots, t_{r-1} < t$. Thus, $b_{r-1} = j$. That is, $\alpha_{b_{r-1}}[t_{r-1}] = Y_j[t_{r-1}]$. As $t_{r-1} < t$ and $j \notin Z$ we have $\alpha_{b_{r-1}}[t_{r-1}] = Y_j[t_{r-1}] \to \alpha_{b_r}[t] = Y_j[t]$ as a sub-trail of $trlG'_Z$. Clearly, $y_j[t_{r-1}]$ is not a collider. As $t_{r-1} < t$, we have $y_j[t_{r-1}] \in \theta$. Thus the trail cannot be active.

Recall the definitions of s_k and τ_{k+1} during construction of the trail in G. If condition P2) is violated, then a sub-path of $trlG$, $v_{m-1} \to v_m \leftarrow v_{m+1}$, must have a collider, v_m, such that $v_m \notin Z$ and $v_{m+1} \notin Z$ where $m = \{2, 3 \cdots, k-1\}$. If $v_{m+1} = j$ and $\tau_{m+1} = t$, P1) also fails, and the argument above shows that the trail in G'_Z is not active. If $v_{m+1} = j$ and $\tau_{m+1} < t$ then we have that $\alpha_{v_{m+1}}[\tau_{m+1}] = Y_{v_{m+1}}[\tau_{m+1}] \in \theta$ which is an observed node along the trail and is not a collider. Thus, the trail $trlG'_Z$ cannot be

active. So, assume that $v_{m+1} \neq j$. By condition $loop_i$, $m+1 \neq i$. As $v_m \leftarrow v_{m+1} \in trlG$, by construction we must have $Y_{v_m}[s_m] = \alpha_{v_m}[s_m] \leftarrow \alpha_{v_{m+1}}[\tau_{m+1}] = Y_{v_{m+1}}[\tau_{m+1}]$ along $trlG'_Z$ with $\tau_{m+1} < s_m < t$. Note that since $v_{m+1} \notin Z$ and $\tau_{m+1} < t$, $\alpha_{v_{m+1}}[\tau_{m+1}] = Y_{v_{m+1}}[\tau_{m+1}]$ is an observed non-collider in θ. Thus, the trail cannot be active.

Finally consider the case that P3) is violated. Then along the trail, $trlG$, in G, there must be a sub-trail $v_{m-1} - v_m - v_{m+1}$ such that the intermediate node, v_m, is not a collider and $v_m \notin Z$. As v_m is not a collider, there is one outgoing directed edge from v_m in the trail $trlG$ to either v_{m-1} or v_{m+1}. By construction, there must be a corresponding node $\alpha_{v_m}[t_f]$ in the trail $trlG'_Z$ such that it has an outgoing edge to either $\alpha_{v_{m-1}}[t_p]$ or $\alpha_{v_{m+1}}[t_q]$ for some $t_p > t_m$ or $t_q > t_m$ respectively. Clearly, there is one $\alpha_{v_m}[t_m]$ in $trlG'_Z$ which is a non-collider. Then, as $v_m \notin Z$, we must have that $\alpha_{v_m}[t_m] = W_{v_m}[t_m] = Y_{v_m}[t_m]$. Note that $v_m \neq i$ by condition $loop_i$. As $t_m < t$, $\alpha_{v_m}[t_m]$ is an intermediate non-collider node in θ and is thus observed. Hence, $trlG'_Z$ cannot be active.

C.2 Proof for Theorem 7

Suppose $i \to j$ is in A_Z. Then there is a trail, trl_G, described by $i = v_1 - v_2 - \cdots - v_k = j$ in G satisfying conditions in Definition 26. We will first construct a trail in the perturbed DBN, G'_Z, from a node in $W_i^{(t-1)}$ to $W_j[t]$ for some $t > 0$. We can construct a trail in G'_Z as follows: for all $l \in \{1, 2, \ldots, k-1\}$, set $t_l = t_{l+1} - k_{v_{l+1}v_l}$ if $v_l \to v_{l+1}$ holds in $trlG$. Otherwise, set $t_l = t_{l+1} + k_{v_l v_{l+1}}$ if $v_l \leftarrow v_{l+1}$ holds in $trlG$. Such a construction is feasible because by condition C 1), numbers $k_{v_{l+1}v_l}$ and $k_{v_l v_{l+1}}$ exists for all $l \in \{1, 2, \ldots, k-1\}$ and at all times. Thus, we have a trail $Y_i[t_1] - Y_{v_2}[t_2] - Y_{v_3}[t_3] - \ldots - Y_{v_{k-1}}[t_{k-1}] - Y_j[t_k]$. For all $m \in \{1, 2, \ldots, k\}$ if $v_m \in Z$, there exists a number $k_m > 0$ following conditions C 2). If B 2) also holds, then $k_m \geq 0$. Let $t > \max\{t_1, \ldots, t_{k-1}\}$, and for all $m \in \{1, 2, \ldots, k\}$ if $v_m \in Z$, let $t > t_m + k_m$ also hold. Depending on whether i or j is a perturbed node, we have four cases on either end of the above trail.

A) Consider the case $i, j \in Z$. As $i \in Z$, using condition C 2) $U_i[t_1 + k_i] \leftarrow Y_i[t_1]$ holds true. Choose t sufficiently large so that $t > t_1 + k_i$ also holds. As $j \in Z$, using C 2), t can be sufficiently large so that we have $Y_j[t_k] \to U_j[t]$ where $t = t_k + k_j$ and $k_j \geq 1$. If B 1) holds, then we can choose t sufficiently large such that at the

end of the trail we take s steps from $Y_j[t_k]$ to $U_j[t]$ such that the tail is of the form $Y_j[t_k] \to Y_j[t_k + k'_j] \to \cdots \to Y_j[t_k + sk'_j] \to U_j[t]$ with $t = t_k + sk'_j + k_j$. Thus, the constructed trail in G'_Z is either $W_i[t_1 + k_i] = U_i[t_1 + k_i] \leftarrow Y_i[t_1] - Y_{v_2}[t_2] - Y_{v_3}[t_3] - \cdots - Y_{v_{k-1}}[t_{k-1}] - Y_j[t_k] \to U_j[t] = W_j[t]$, or $W_i[t_1 + k_i] = U_i[t_1 + k_i] \leftarrow Y_i[t_1] - Y_{v_2}[t_2] - Y_{v_3}[t_3] - \cdots - Y_{v_{k-1}}[t_{k-1}] - Y_j[t_k] \to Y_j[t_k + k'_j] \to \cdots \to Y_j[t_k + sk'_j] \to U_j[t] = W_j[t]$ with $t > \max\{t_1 + k_i, t_1, \ldots, t_k, \ldots, t_k + sk'_j\}$, and for all $m \in \{1, 2, \ldots, k\}$ if $v_m \in Z$, $t > t_m + k_m$.

B) Consider the case $i \in Z$ but $j \notin Z$. Choose t as t_k. As $i \in Z$, using condition C 2) $U_i[t_1 + k_i] \leftarrow Y_i[t_1]$ holds true. Choose t sufficiently large so that $t > t_1 + k_i$ also holds. Thus, we have constructed a trail in G'_Z which is of the form: $W_i[t_1 + k_i] = U_i[t_1 + k_i] \leftarrow Y_i[t_1] - Y_{v_2}[t_2] - Y_{v_3}[t_3] - \cdots - Y_{v_{k-1}}[t_{k-1}] - Y_j[t] = W_j[t]$ with $t > \max\{t_1 + k_i, t_1, \ldots, t_{k-1}\}$, and for all $m \in \{1, 2, \ldots, k\}$ if $v_m \in Z$, $t > t_m + k_m$.

C) Consider the case $i \notin Z$ but $j \in Z$. Following arguments presented in case (A) we conclude that the constructed trail of form $W_i[t_1] = Y_i[t_1] - Y_{v_2}[t_2] - Y_{v_3}[t_3] - \cdots - Y_{v_{k-1}}[t_{k-1}] - Y_j[t_k] \to U_j[t] = W_j[t]$, or of form $W_i[t_1] = Y_i[t_1] - Y_{v_2}[t_2] - Y_{v_3}[t_3] - \cdots - Y_{v_{k-1}}[t_{k-1}] - Y_j[t_k] \to Y_j[t_k + k'_j] \to \cdots \to Y_j[t_k + sk'_j] \to U_j[t] = W_j[t]$ exists in the perturbed DBN G'_Z with $t > \max\{t_1, \ldots, t_k, \ldots, t_k + sk'_j\}$, and for all $m \in \{1, 2, \ldots, k\}$ if $v_m \in Z$, $t > t_m + k_m$.

D) Consider the case $i \notin Z$ and $j \notin Z$. Following arguments presented in Case (B) we conclude that the trail $W_i[t_1] = Y_i[t_1] - Y_{v_2}[t_2] - Y_{v_3}[t_3] - \cdots - Y_{v_{k-1}}[t_{k-1}] - Y_j[t] = W_j[t]$ exists in the perturbed DBN G'_Z with $t > \max\{t_1, \ldots, t_{k-1}\}$, and for all $m \in \{1, 2, \ldots, k\}$ if $v_m \in Z$, $t > t_m + k_m$.

We will now argue that in each of the cases above, the constructed trail is active given $\theta := \{W_j^{(t-1)}, \mathcal{W}_{\overline{ji}}^{(t-1)}\}$.

Sub-trails with colliders: For all the trails in G'_Z constructed under various cases above consider a sub-trail of the form $Y_{v_{m-1}}[t_{m-1}] \to Y_{v_m}[t_m] \leftarrow Y_{v_{m+1}}[t_{m+1}]$. Clearly, v_m cannot be either i or j. If $v_m \notin Z$ then as $t_m < t$, we have $Y_{v_m}[t_m] \in W_{\overline{ji}}^{(t-1)}$ and thus the sub-trail is active. If $v_m \in Z$ then the corrupted version of $Y_{v_m}[t_m]$ is $U_{v_m}[t_m + k_{v_m}] = W_{v_m}[t_m + k_{v_m}]$ and as $t_m + k_{v_m} < t$, we have $W_{v_m}[t_m + k_m] \in W_{\overline{ji}}^{(t-1)}$. Thus the collider $Y_{v_m}[t_m]$ has a descendant $W_{v_m}[t_m + k_{v_m}] \in \theta$. Thus the sub-trail remains active. Thus no collider can deactivate the trails in G'_Z.

Sub-trails with with no colliders: Now consider any node $Y_{v_m}[t_m]$ which is not a collider. Note that in the trails for the cases (A), (B), (C), and (D), Y_j and Y_i can only appear as an intermediate node only if they are corrupted. In such cases, neither $Y_i[t_1]$ nor $Y_j[t_k]$ belong to θ. Thus, if Y_j or Y_i are intermediate nodes, they cannot deactivate the trails given θ. Consider an intermediate node $v_m \notin \{i, j\}$. From Definition 26P 3), v_m is corrupted. Thus $Y_{v_m}[t_m] \neq W_{v_m}[t_m]$ and $Y_{v_m}[t_m]$ cannot deactivate the trail as $Y_{v_m}[t_m] \notin \theta$. □

Appendix D

Estimation of Conditional Directed Information

D.1 Proof for Theorem 8

To prove the theorem, we require two results from [182]. The following lemma shows that with sufficiently large data, the conditional probability assignment by CTW converges to the true probability assignment for a Markov process.

Lemma 9. *Let Q be the probability assignment described by CTW. Let X be a stationary and finite alphabet Markov process with finite Markov order which is bounded by the prescribed tree depth of CTW algorithm. Let P be the true probability for X. Then,*

$$\lim_{n\to\infty} Q(x[n] \mid x^{(n-1)}) - P(x[n] \mid x^{(n-1)}) = 0 \ \textit{P-as.} \tag{D.1}$$

Next, we will later use the following proposition which is a rephrased result from [182].

Proposition 3. *Let Q be the probability assignment in the CTW algorithm. Suppose, X, Y are jointly stationary irreducible aperiodic finite-alphabet Markov processes whose order is bounded by the prescribed tree depth of the CTW algorithm. Let $\hat{H}(y^{(n)} \parallel x^{(n)}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid x^{(i-1)}, y^{(i-1)}) \cdot \log \frac{1}{Q(y[i]|x^{(i-1)},y^{(i-1)})}$. Then,*

$$\lim_{n\to\infty} \hat{H}(y^{(n)} \parallel x^{(n)}) - H_r(Y \parallel X) = 0 \qquad \textit{P-a.s,} \tag{D.2}$$

Recall the expression for conditional DI estimator from (5.9):

$$\hat{I}(x^{(n)} \to y^{(n)} \parallel z^{(n)}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)}) \cdot \log \frac{1}{Q(y[i] \mid y^{(i-1)}, z^{(i-1)})} - \frac{1}{n}\sum_{i=1}^{n}\sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)}) \cdot \log \frac{1}{Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)})} \quad \text{(D.3)}$$

We will show that the first term (call it T1) in equation (D.3) converges to $H_r(Y \parallel Z)$ and the second term (call it T2) in (D.3) converges to $H_r(Y \parallel X, Z)$.

Convergence of T2: Let $V = \{X, Z\}$. Thus, T2 can be written as $\hat{H}(y^{(n)} \parallel v^{(n)}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid v^{(i-1)}, y^{(i-1)}) \cdot \log \frac{1}{Q(y[i]|v^{(i-1)}, y^{(i-1)})}$. Using, proposition 3, we thus have that $\lim_{n\to\infty} \hat{H}(y^{(n)} \parallel v^{(n)}) \to H_r(Y \parallel V)$ almost surely.

Convergence of T1: Subtract $H_r(Y \parallel Z)$ from $T1$ and express $T1 - H_r(Y \parallel Z) = F_n + S_n$ where,

$$F_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{y[i]\in\mathcal{Y}} P(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)}) \cdot \log P(y[i] \mid y^{(i-1)}, z^{(i-1)}) - \frac{1}{n}\sum_{i=1}^{n}\sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)}) \cdot \log Q(y[i] \mid y^{(i-1)}, z^{(i-1)}), \quad \text{(D.4)}$$

$$S_n = -\frac{1}{n}\sum_{i=1}^{n}\sum_{y[i]\in\mathcal{Y}} P(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)}) \cdot \log P(y[i] \mid y^{(i-1)}, z^{(i-1)}) - H_r(Y \parallel Z) \quad \text{(D.5)}$$

By ergodicity, S_n converges to zero almost surely. We need to show that F_n converges to zero almost surely. Rewrite $F_n = \frac{1}{n}\sum_{i=1}^{n}\beta_i$ where,

$$\beta_i = \sum_{y[i]\in\mathcal{Y}} P(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)}) \cdot \log P(y[i] \mid y^{(i-1)}, z^{(i-1)}) - \sum_{y[i]\in\mathcal{Y}} Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)}) \cdot \log Q(y[i] \mid y^{(i-1)}, z^{(i-1)}) \quad \text{(D.6)}$$

By Lemma 9, the CTW probabilities $Q(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)})$ converges to true probabilities $P(y[i] \mid x^{(i-1)}, y^{(i-1)}, z^{(i-1)})$ almost surely. Therefore,

$$\lim_{i\to\infty} \beta_i = 0 \qquad \text{P-a.s.} \tag{D.7}$$

Hence, by Cesaro mean [193] we have:

$$\lim_{n\to\infty} F_n = \lim_{n\to\infty} \frac{1}{n}\beta_i = 0 \quad \text{P-a.s.} \quad \square \tag{D.8}$$

Appendix E

Corruption Detection in Networks of Bi-directional Dynamical Systems

E.1 Proof of Proposition 1

Let $\hat{A}_Z$ be the edge set described in the proposition. We will show that $\hat{A}_Z = A_Z$.

First, we show that $\hat{A}_Z \subset A_Z$. Suppose, $i \to j \in \hat{A}_Z$. If $i \rightleftharpoons j$ was an edge of G, then $i \to j \in A_Z$.

Now, consider the case of a trail of length at least 3 with $i = v_1 \rightleftharpoons v_2 \rightleftharpoons v_3 \rightleftharpoons \cdots \rightleftharpoons v_k = j$.

Since the network is bi-directional, we can choose the directionality of the edges. We will show that by suitable choice of directionality, we can retrieve a directed path between i and j that satisfy all the conditions P1), P2) and P3) for the link $i \to j$ to be in A_Z. So, for each pair of nodes along the trail, set the directionality as follows:

- For $m \geq 2$, if $v_m \notin Z$, set $v_{m-1} \to v_m \leftarrow v_{m+1}$.
- Set all other edges in the $\to$ direction.

Since no two consecutive nodes, v_m and v_{m+1} with $m \geq 2$ are unperturbed, this construction is feasible.

P1) If $v_k = j \notin Z$, then, since the trail has $k \geq 3$, we must have $k-1 \geq 2$. So, we must have that $v_{k-1} \in Z$. Thus, by our convention for choosing directions, we have $v_{k-1} \to v_k$. Thus, P1) holds.

P2) If v_m is a collider, then we have $v_m \leftarrow v_{m+1}$. This directionality is chosen only when when $v_m \notin Z$. Furthermore, since v_m is a collider, we must have $m \geq 2$. Thus, we have $v_{m+1} \in Z$. Therefore, P2) holds.

P3) If v_m is an intermediate node which is not a collider, then by construction, it cannot be unperturbed. Thus, P3) holds.

Now, we show that $A_Z \subset \hat{A}_Z$. Say that $i \to j \in A_Z$. Let $i = v_1 - \cdots - v_k = j$ be trail in G that satisfies conditions P1), P2), and P3). If the trail is $i - j$, then from B1) we have $i \to j \in \hat{A}_Z$.

Now consider the case that the trail has length of at least 3. We must show that for any pair v_m, v_{m+1} with $m \geq 2$, at least one of the nodes is in Z. Since $m \geq 2$, we must have $v_{m-1} - v_m - v_{m+1}$ on the trail. If m is a collider, then P2) implies that $v_{m+1} \in Z$. If m is not a collider, then P3) implies that $v_m \in Z$. Thus, at least one of v_m or v_{m+1} is perturbed. Thus, $i \to j \in \hat{A}_Z$.

E.2 Proof for Theorem 9

Let statements T1) and T2) be true. Now, we will show that j is corrupted if and only if statement T3) holds.

($\Rightarrow$) Say that j is corrupted. Let p and q be bi-directional neighbors of j in G_Z. We will show that $p \to q$ in G_Z as well. Since p and q are arbitrary, switching their roles shows that we must have $p \rightleftharpoons q$. Furthermore, since p and q are arbitrary bidirectional neighbors of j, $\{j\} \cup \text{bidNr}(j)$ must form a bidirectional clique in G_Z.

Now, we will show that $p \to q \in A_Z$. Since $p \rightleftharpoons j \in A_Z$ and $j \rightleftharpoons q \in A_Z$, by Proposition 1, there must be a trail in G of the form

$$p = v_1 \rightleftharpoons \cdots \rightleftharpoons v_k = j = w_1 \rightleftharpoons \cdots \rightleftharpoons w_l = q \tag{E.1}$$

such that the following conditions hold:

- If $v_m \rightleftharpoons v_{m+1}$ with $m \geq 2$, then at least one of v_m or v_{m+1} must be perturbed.
- If $w_m \rightleftharpoons w_{m+1}$ with $m \geq 2$, then at least one of w_m or w_{m+1} must be perturbed.

Note that these conditions automatically hold if the paths have length 2.

Since $j \in Z$, at least one node is perturbed on each of $v_{k-1} \rightleftharpoons j$ and $j \rightleftharpoons w_2$. Thus, there are no consecutive unperturbed nodes along the subtrail:

$$v_2 \rightleftharpoons \cdots \rightleftharpoons v_{k-1} \rightleftharpoons j \rightleftharpoons w_2 \cdots \rightleftharpoons w_l = q \tag{E.2}$$

Thus, $p \to q$.

($\Leftarrow$) As $i \to j \in A_Z$ and $j \to i \notin A_Z$, $i \to j$ is a spurious link. By Proposition 1, there exists a trail in G, $trl_G : i = v_1 \rightleftharpoons v_2 \rightleftharpoons v_3 \rightleftharpoons \cdots \rightleftharpoons v_{k-1} \rightleftharpoons v_k = j$, such that there no two consecutive unperturbed nodes along the subtrail $v_3 \rightleftharpoons v_4 \rightleftharpoons \cdots \rightleftharpoons v_{k-1} \rightleftharpoons v_k = j$. Suppose, $j \notin Z$ for the sake of contradiction. We will show that T3) does not hold, leading to a contradiction. That is, we will show that there exists at least one subset of nodes $S \subset \text{bidNr}(j)$, such that S does not form a bi-directional clique with j in G_Z. Since $j \notin Z$, by Assumption 3, there is at least one unperturbed node l such that $j \rightleftharpoons l$ holds in G. Therefore, $v_{k-1} \rightleftharpoons j \rightleftharpoons l$ exists in G. That is, v_{k-1} and l are bi-directional neighbors of j. We will now prove that $\{v_{k-1}, l\}$ does not form a bi-directional clique with j in G_Z. Particularly, we will show that $v_{k-1} \to l \notin A_Z$.

To show that $v_{k-1} \to l \notin A_Z$, we will show that there exists no trail connecting v_{k-1} and l in G satisfying conditions P1), P2) and P3). Suppose, there exists a trail different from $v_{k-1} \rightleftharpoons j \rightleftharpoons l$ in G. Such a trail will imply two paths connecting v_{k-1} and l in the topology, G^τ, of G. This contradicts the assumption that G^τ is a tree. Therefore, $v_{k-1} \rightleftharpoons j \rightleftharpoons l$ is the only trail in G that connects v_{k-1} and l. Now, we will show that all directed paths in this trail violate at least one of P1), P2) and P3). Consider the directed path $v_{k-1} \to j \leftarrow l$ where j is a collider. As $j \notin Z$ and $l \notin Z$, condition P2) is not met. Now consider $v_{k-1} \to j \to l$. Then, as $j \notin Z$, P3) is violated. Consider the directed path $v_{k-1} \leftarrow j \to l$. As $j \notin Z$, P3) is violated. Finally consider $v_{k-1} \leftarrow j \leftarrow l$. Again, as $j \notin Z$, P3) is violated. Thus all possible directed paths in $v_{k-1} \rightleftharpoons j \rightleftharpoons l$ violate atleast one of P1), P2) and P3). Thus $v_{k-1} \to l$ cannot be in the perturbed graph. Therefore, $\{v_{k-1}, l, j\}$ does not form a bi-directional clique in G_Z. This completes the proof that j must be a perturbed node.

Appendix F

Topology Learning in Radial Dynamical Systems with Unreliable Data

F.1 Proof of Lemma 2

($\Rightarrow$) We will show that if i is neither a leaf node nor a corrupt node, then $\mathcal{N}_u(i) \cup \{i\}$ does not form a clique in G^U. Note that all nodes in $1-hop(i)$ and all nodes in $2-hop(i)$ will be neighbors of i in moral graph and hence are neighbors of i in G^U. We will show that there exists a pair of nodes $a, b \in \mathcal{N}_u(i)$ such that $a - b$ does not hold true in G^U and thus $\mathcal{N}(i)_u \cup \{i\}$ cannot form a clique in G^U.

By Assumption C2), we have that there is at least one 1-hop neighbor and one 2 -hop neighbor for every node in G^T. Consider $a \in 1 - hop(i)$ and $b \in 2 - hop(i)$. Then, a path $a - i - p - b$ exists in G^T for some node $p \in V$. Either a is a leaf node or not in G^T. Suppose, a is a leaf node. As G^T is a tree, the path $a - i - p - b$ is unique. Thus, all possible paths between a and b in G^M goes through at least one of i, p. As a is a leaf node, by condition C 1), i, p are not corrupt. Thus, $a - b \notin A^U$.

Suppose a is not a leaf node. Then, there exists a path $c - a - i - p - b$ in G^T. Thus, $c \in 2 - hop(i)$ and a neighbor of i in the perturbed graph, G^U. We show that c, b despite being neighbors of i in G^U, $c - b$ does not hold true in G^U. Similar to argument

above, since G^T is a tree, the path connecting c and b, $c-a-i-p-b$ is unique in G^T. Thus, all possible paths between c and b in G^M goes through at least one of a, p and i. By condition C2), both a and p cannot be corrupt. Thus all possible paths between c and b in G^M goes through at least one non-perturbed node. Thus, $c-b \notin A^U$.

($\Leftarrow$) Suppose i is a leaf node in G^T or a corrupt node. We will show that $\mathcal{N}_u(i) \cup \{i\}$ forms a clique in the perturbed kin graph, G^U. Using Proposition 2, $\mathcal{N}_u(i) \cup \{i\} = 1-hop(i) \cup 2-hop(i) \cup \{i\}$.

i is a leaf node: There is only one non-leaf node nl which is a neighbor of i in G^T. Any pair of 2-hop neighbors of i in G^T, k_1, k_2, has a common parent of nl in the generative graph. Thus, $k_1 - nl - k_2$ holds in G^T and neighbors in G^M. Hence, $k_1 - k_2$ holds in G^U. Thus, $\mathcal{N}_u(i) \cup \{i\}$ forms a clique in G^U.

i is a corrupt node: For any k_1, $k_2 \in \mathcal{N}_u(i)$, there is a path $k_1 - i - k_2$ in the moral graph G^M. As i is a corrupt node, $k_1 - k_2$ holds in G^U. Thus, $\mathcal{N}_u(i) \cup \{i\}$ forms a clique in G^U. □

F.2 Proof of Theorem 10

Structure of inverse power spectra due to corruption

The expressions for $\Phi_{uu}^{-1}(\omega)$ are explicitly required for the results. To this, we will begin by describing the general structure of $\Phi_{uu}^{-1}(\omega)$. For compact notation, we will often drop the ω arguments.

Consider a network of N nodes. For $p = 1, \ldots, N$ if p is not a perturbed node, set $h_p(\omega) = 1$ and $\theta_p(\omega) = 0$. With this notation, (2.10) implies that the entries of Φ_{uu} are given by:

$$(\Phi_{uu})_{pq} = \begin{cases} h_p(\Phi_{xx})_{pq} h_q^* & \text{if } p \neq q \\ h_p(\Phi_{xx})_{pp} h_p^* + \theta_p & \text{if } p = q. \end{cases}$$

When $p \neq q$, there is no θ term because the perturbations were assumed to be independent.

In matrix notation, we have that:

$$\Phi_{uu} = H \Phi_{xx} H^* + \sum_{k=1}^{n} \Theta_{v_k}$$

where H is the diagonal matrix with entries h_p on the diagonal and $\Theta_{v_k}(\omega) = b_{v_k}\theta_{v_k}(\omega)b_{v_k}^T$ where b_{v_k} is the canonical unit vector with 1 at entry v_k.

For $k = 0, \ldots, n-1$ set $\Psi_k = H\Phi_{xx}H^* + \sum_{m=1}^{k}\Theta_{v_m}$. Here, $\Phi_{uu}^{-1} = \Psi_n^{-1}$. We can inductively define these matrices as:

$$\Psi_0 = H\Phi_{xx}H^*, \Psi_{k+1} = \Psi_k + b_{v_{k+1}}\theta_{v_{k+1}}b_{v_{k+1}}^T. \tag{F.1}$$

Note that $\Psi_0^{-1}(i,j)$ can be expressed as follows:

$$\Psi_0^{-1}(i,j) = \frac{1}{\overline{h_i}(\omega)}\Phi_{xx}^{-1}(i,j)\frac{1}{h_j(\omega)}. \tag{F.2}$$

Combining Woodbury matrix identity in (F.1) implies that

$$\Psi_{k+1}^{-1} = \Psi_k^{-1} - \Gamma_{k+1}, \tag{F.3}$$

where $\Gamma_{k+1} := \Psi_k^{-1}b_{v_{k+1}}b_{v_{k+1}}^T\Psi_k^{-1}\Delta_{k+1}^{-1}$, and $\Delta_{k+1} = \theta_{v_{k+1}}^{-1} + \Psi_k^{-1}(v_{k+1}, v_{k+1})$ being a scalar.

Expression for (i,j) entry in $\Phi_{uu}^{-1}(\omega)$

Now, we will show that for each candidate node $i \in B$, $\Phi_{uu}^{-1}(i,j) = \Psi_1^{-1}(i,j)$. For each $i \in B$, we will inductively prove that for $k = 2, \ldots, n$, $\Psi_k^{-1}(i,j) = \Psi_1^{-1}(i,j)$. Consider case $k = 2$. Using (F.3), $\Psi_2^{-1}(i,j) = \Psi_1^{-1}(i,j) - \Gamma_2(i,j)$. Note that $\Gamma_2(i,j) = \Psi_1^{-1}(i,v_2)\Psi_1^{-1}(v_2,j)\Delta_2^{-1}$. Similarly, using (F.3), $\Psi_1^{-1}(i,v_2) = \Psi_0^{-1}(i,v_2) - \Gamma_1(i,v_2)$ where $\Gamma_1(i,v_2) := \Psi_0^{-1}(i,v_1)\Psi_0^{-1}(v_1,v_2)\Delta_1^{-1}$. Here, if i is a corrupt node, $v_1 = i$. By (F.2), $\Psi_0^{-1}(v_1,v_2) = \frac{\Phi_{xx}^{-1}(v_1,v_2)}{\overline{h_{v_1}}h_{v_2}}$. As v_1 and v_2 are at least 3-hops away in G^T, using (7.6), $\Phi_{xx}^{-1}(v_1,v_2) = 0$. Thus, $\Gamma_1(i,v_2) = 0$. Again, by (F.2) $\Psi_0^{-1}(i,v_2) = \frac{\Phi_{xx}^{-1}(i,v_2)}{\overline{h_i}h_{v_2}}$. Since i is either a leaf node or a corrupt node, we have that i, v_2 are at least 3 hops away from each other. Using this and (7.6) we have that $\Phi_{xx}^{-1}(i,v_2) = 0$. This implies $\Psi_0^{-1}(i,v_2) = 0$. Therefore, $\Psi_1^{-1}(i,v_2) = 0$ and hence $\Gamma_2(i,j) = 0$. Thus we have proved that $\Psi_2^{-1}(i,j) = \Psi_1^{-1}(i,j)$.

Now assume that the claim holds for some $k > 2$. That is, $\Psi_k^{-1}(i,j) = \Psi_1^{-1}(i,j)$. By (F.3), $\Psi_{k+1}^{-1}(i,j) = \Psi_k^{-1}(i,j) - \Gamma_{k+1}(i,j)$, where $\Gamma_{k+1}(i,j) = \Psi_k^{-1}(i,v_{k+1})\Psi_k^{-1}(v_{k+1},j)\Delta_{k+1}^{-1}$. Using the induction hypothesis, $\Psi_k^{-1}(i,v_{k+1}) = \Psi_1^{-1}(i,v_{k+1})$. As v_1, v_{k+1} and i, v_{k+1} are at least 3 hops away from each other respectively, using the same argument as described before for v_1, v_2 and i, v_2, we have $\Gamma_{k+1}(i,v_{k+1}) = 0$ and $\Psi_{k+1}^{-1}(i,j) = \Psi_k^{-1}(i,j) = \Psi_1^{-1}(i,j)$. As $\Phi_{uu}^{-1} = \Psi_k^{-1}$ for $k = n$, we have established that $\Phi_{uu}^{-1}(i,j) = \Psi_1^{-1}(i,j)$.

Proof of Theorem 10

($\Leftarrow$) We will show that if $i \in B$ is a leaf node in G^T, then there is at most only one node $j \in \mathcal{N}_u(i)$ such that $\angle\Phi_{uu}^{-1}(i,j)(\omega)$ is non-constant for all $\omega \in (-\pi, \pi]$. By Proposition 2, $\mathcal{N}_u(i) = hop_1(i) \cup hop_2(i)$. Moreover as i is a leaf node, any node $j \in \mathcal{N}_u(i)$ is not a corrupt node. Therefore, using (F.2) and preceding discussion, we have

$$\Phi_{uu}^{-1}(i,j) = \Psi_1^{-1}(i,j) = \Psi_0^{-1}(i,j) = \Phi_{xx}^{-1}(i,j). \tag{F.4}$$

Since i is a leaf node, there is only node in $hop_1(i)$. Suppose r is that node. We will show that for all $j \neq r \in \mathcal{N}_u(i)$ (this means $j \in hop_2(i)$), $\angle\Phi_{uu}^{-1}(i,j)$ is a constant either 0 or π while $\angle\Phi_{uu}^{-1}(i,r)$ is non-constant for all ω.

Take any $j \in hop_2(i)$. Let $q \in V$ be the common neighbor of i and j in G^T. Combining (F.4) and (7.6) we have:

$$\Phi_{uu}^{-1}(i,j) = \overline{\mathcal{G}}_{qi}(\omega)\mathcal{G}_{qj}(\omega)\Phi_{e_q}^{-1} \tag{F.5}$$

$$= \frac{b_{qi}b_{qj}\Phi_{w_q}^{-1}}{|S_q|^4}, \tag{F.6}$$

which is a real scalar. Thus, $\angle\Phi_{uu}^{-1}(i,j) = 0$ or π.

Now, consider the node r. By (F.4) and (7.6) we have:

$$\Phi_{uu}^{-1}(i,r) = -\frac{b_{ir}\Phi_{w_i}^{-1}}{S_i(\omega)|S_i(\omega)|^2} - \frac{b_{ri}\Phi_{w_r}^{-1}}{S_r(\omega)|S_r|^2}. \tag{F.7}$$

Recall that $S_i(\omega)$ and $S_r(\omega)$ are transfer functions obtained from the differential equation in (7.1) and are not real scalars for all ω. Thus $\Phi_{uu}^{-1}(i,r)$ has a non-constant phase response.

($\Rightarrow$) We will show that if $i \in B$ is a corrupt node in G^T, then there are at least two neighbors of i in G^U such that the corresponding entries in the inverse PSD have non-constant transfer functions. By Assumption 2 on the location of corrupt nodes, every corrupt node has at least two 1 hop neighbors in G^T. By construction of perttubed graph, G^U, these nodes will also be neighbors of i in G^U. Say these nodes are p, r. We will now show that $\Phi_{uu}^{-1}(i,p)$, $\Phi_{uu}^{-1}(i,r)$ are non-constant transfer functions. Consider $\Phi_{uu}^{-1}(i,p)$.

Now, $\Phi_{uu}^{-1}(i,p) = \Psi_0^{-1}(i,p) - \Psi_0^{-1}(i,i)\Psi_0^{-1}(i,p)\Delta_1^{-1}$. By (F.2), $\Psi_0^{-1}(i,p) = \frac{\Phi_{xx}^{-1}(i,p)}{\overline{h_i}h_p}$. Then,

$$\Phi_{uu}^{-1}(i,p) = \frac{\Phi_{xx}^{-1}(i,p)}{\overline{h_i}} - \frac{\Delta_1^{-1}\Phi_{xx}^{-1}(i,i)\Phi_{xx}^{-1}(i,p)}{\overline{h_i}|h_i|^2}. \tag{F.8}$$

Using (7.6) we have:

$$\Phi_{xx}^{-1}(i,p) = -\frac{b_{ip}\Phi_{w_i}^{-1}}{S_i(\omega)|S_i(\omega)|^2} - \frac{b_{pi}\Phi_{w_p}^{-1}}{S_p(\omega)|S_p|^2}. \tag{F.9}$$

Recall that $S_i(\omega)$ and $S_p(\omega)$ are transfer functions obtained from the differential equation in (7.1) and are not real scalars for all ω. Therefore, by (F.8) and (F.9) we can see that $\angle\Phi_{uu}^{-1}(i,p)$ is not a constant for all ω. The case for $\angle\Phi_{uu}^{-1}(i,r)$ being non-constant can be shown similarly. □

F.3 Proof of Theorem 11

Following the discussion in Section 7.3.2, the only result that needs to be proved is Lemma 5, corresponding to the *Placement of Corrupt Nodes*.

Proof. Since $\{p,q,l,r,s\}$ forms a clique in G^U and G^T is a tree, it follows that l is located at the point of disconnection between $p-q$ and $r-s$. What needs to be shown is the correct alignment among the paths $q-p-l-r-s$, $p-q-l-s-r$ and $q-p-l-s-r$ and $p-q-l-r-s$ in G^T. To this we will analyze the phase of inverse PSD entry corresponding to pairs from $\{p,q\}\times\{r,s\}$ described in Proposition 5. Before that we will need the following proposition that gives an expression for inverse PSD entry corresponding to pairs from $\{p,q\}\times\{r,s\}$.

Proposition 4. *Suppose $p-q-l-r-s$ holds in G^T where l is a corrupt node. Then, for any $a\in\{p,q\}$ and $b\in\{r,s\}$, $\Phi_{uu}^{-1}(\omega)(a,b) = \Psi_1^{-1}(\omega)(a,b)$, where Ψ_1^{-1} is defined by* (F.3) *and $v_1 = l$.*

Proof. For $k = 2,3,\ldots,n$ perturbed nodes we will inductively show that $\Psi_k^{-1}(a,b) = \Psi_1^{-1}(a,b)$.

We will require the following claim: for any v_k being a perturbed node ($k>1$), at the most only one of $a-v_k$ or v_k-b holds in G^T. Note that there is already a path, $p-q-l-r-s$, consisting of a and b. As $v_k\neq l$ and G^T is a tree, existence of $a-v_k-b$

violates the assumption that G^T is a tree. This proves the claim. In other words, at least one of $a - v_k$ or $v_k - b$ does not hold in G^T. Suppose $v_k - b$ does not hold true. (The case $a - v_k$ can be shown similarly). Refer to this result as $R1$.

Consider $k = 2$. We will show that $\Psi_2^{-1}(a,b) = \Psi_1^{-1}(a,b)$. Using (F.3), $\Psi_2^{-1}(a,b) = \Psi_1^{-1}(a,b) - \Gamma_2(a,b)$. Note that $\Gamma_2(a,b) = \Psi_1^{-1}(a,v_2)\Psi_1^{-1}(v_2,b)\Delta_2^{-1}$. We will show that $\Gamma_2(a,b) = 0$. Using (F.3), $\Psi_1^{-1}(v_2,b) = \Psi_0^{-1}(v_2,b) - \Gamma_1(v_2,b)$ where $\Gamma_1(v_2,b) := \Psi_0^{-1}(v_2,v_1)\Psi_0^{-1}(v_1,b)\Delta_1^{-1}$. By (F.2), $\Psi_0^{-1}(v_1,v_2) = \frac{\Phi_{xx}^{-1}(v_1,v_2)}{h_{v_1}h_{v_2}}$. As v_1 and v_2 are at least 3-hops away in G^T, using (7.6), $\Phi_{xx}^{-1}(v_1,v_2) = 0$. Thus, $\Gamma_1(v_2,v_1) = 0$. Invoking $R1$, we have that $v_2 - b$ does not hold in G^T. Then, v_2 can either be a 2-hop neighbor of b in G^T or not.

Suppose v_2 is a 2-hop neighbor of b. Then, v_2 cannot be a 1-hop or 2-hop neighbor of a because this leads to two paths connecting a and b: one through v_2 and the other being $p-q-l-r-s$, violating the condition that G^T is a tree. Thus, v_2 is neither a 1-hop neighbor nor a 2-hop neighbor of a in G^T. Thus, using (7.6) we have that $\Phi_{xx}^{-1}(a,v_2) = 0$. By (F.2), $\Psi_0^{-1}(a,v_2) = \frac{\Phi_{xx}^{-1}(a,v_2)}{h_a h_{v_2}}$. This implies $\Psi_0^{-1}(a,v_2) = 0$. Therefore, $\Psi_1^{-1}(a,v_2) = 0$ and hence $\Gamma_2(a,b) = 0$. Thus we have proved that $\Psi_2^{-1}(a,b) = \Psi_1^{-1}(a,b)$.

Now consider that v_2 is not a 2-hop neighbor of b. Then, using (7.6) we have that $\Phi_{xx}^{-1}(v_2,b) = 0$. By (F.2), $\Psi_0^{-1}(v_2,b) = \frac{\Phi_{xx}^{-1}(v_2)}{h_{v_2}h_b}$. This implies $\Psi_0^{-1}(v_2,b) = 0$. Therefore, $\Psi_1^{-1}(v_2,b) = 0$ and hence $\Gamma_2(a,b) = 0$. Thus we have proved that $\Psi_2^{-1}(a,b) = \Psi_1^{-1}(a,b)$.

Now assume that the claim holds for some $k > 2$. That is, $\Psi_k^{-1}(a,b) = \Psi_1^{-1}(a,b)$. Using (F.3), $\Psi_{k+1}^{-1}(a,b) = \Psi_k^{-1}(a,b) - \Gamma_{k+1}(a,b)$. As shown in Theorem 10, $\Psi_k^{-1}(a,v_{k+1}) = \Psi_1^{-1}(a,v_{k+1})$ and $\Psi_k^{-1}(v_{k+1},b) = \Psi_1^{-1}(v_{k+1},b)$. Invoking $R1$, we have that $v_{k+1} - b$ does not hold in G^T. Then, v_{k+1} can either be a 2-hop neighbor of b in G^T or not.

Suppose v_{k+1} is a 2-hop neighbor of b. Similar to argument for v_2, v_{k+1} cannot be a 1-hop or 2-hop neighbor of a. As v_1 and v_{k+1} are at least 3-hops away in G^T, using (7.6), $\Phi_{xx}^{-1}(v_1,v_{k+1}) = 0$. Now consider that v_{k+1} is not a 2-hop neighbor of b. Similar to argument for v_2, $\Psi_1^{-1}(v_{k+1},b) = 0$ and hence $\Psi_k^{-1}(v_{k+1},b) = 0$ and $\Gamma_{k+1}(a,b) = 0$. Thus we have proved that $\Psi_{k+1}^{-1}(a,b) = \Psi_1^{-1}(a,b)$. □

Our next result, Proposition 5, analyzes the phase values of enteries in Φ_{uu}^{-1}.

Proposition 5. *If $p-q-l-r-s$ holds in G^T, then $\angle\Phi_{uu}^{-1}(\omega)(p,s)$ is a constant while $\angle\Phi_{uu}^{-1}(\omega)(p,r)$, $\angle\Phi_{uu}^{-1}(\omega)(q,s)$ and $\angle\Phi_{uu}^{-1}(\omega)(q,r)$ are non-constant for all $\omega \in (-\pi,\pi]$.*

Proof. Using Proposition 4 we have that for any $a \in \{p, q\}$ and $b \in \{r, s\}$, $\Phi_{uu}^{-1}(\omega)(a, b) = \Psi_1^{-1}(\omega)(a, b)$, where Ψ_1^{-1} is defined by (F.3) and $v_1 = l$.

As $v_1 = l$ is the only corrupt node, $\Phi_{uu}^{-1}(\omega)(a, b)$ can be expressed as:

$$\Phi_{uu}^{-1}(a, b) = \Psi_0^{-1}(a, b) - \Psi_0^{-1}(a, l)\Psi_0^{-1}(l, b)\Delta_{v_1}^{-1}. \tag{F.10}$$

Moreover, as a, b are not corrupt nodes, we have that $h_a(\omega) = h_b(\omega) = 1$. Thus, $\Psi_0^{-1}(a, l) = \Phi_{xx}^{-1}(a, l)$ and $\Psi_0^{-1}(l, b) = \Phi_{xx}^{-1}(l, b)$. Now we will show that the term $\Delta_{v_1}^{-1}$ is real valued for all ω. Now, $\Delta_l = d_{v_1}^{-1} + \Psi_0^{-1}(v_1, v_1)$. By (F.2) we have that $\Psi_0^{-1}(v_1, v_1) = \frac{\Phi_{xx}^{-1}(v_1, v_1)}{|h_l(\omega)|^2}$. Using (7.6), we have that $\Phi_{xx}^{-1}(v_1, v_1)$ is real valued and therefore, $\Psi_0^{-1}(v_1, v_1)$ is real valued. As $\theta(\omega)$ is the PSD of autocorrelation of a WSS process it will be real and non-negative valued. (A more formal description is in Chapter 2). Thus, $\Delta_{v_1}^{-1}$ is real valued for all $\omega \in (-\pi, \pi]$.

We now proceed to evaluating $\angle\Phi_{uu}^{-1}(\omega)(a, b)$ for all combinations of $a \in \{p, q\}$ and $b \in \{r, s\}$.

(q, r): In this case q, r are 1-hop neighbors of l in G^T and hence as discussed in Theorem 10, $\Psi_0^{-1}(q, l)$ and $\Psi_0^{-1}(l, r)$ will be non-constant transfer functions. Thus, $\Phi_{uu}^{-1}(q, r)(\omega)$ will be non-constant transfer functions.

(p, r): Here, s is a 1-hop neighbor of l in G^T. Thus as discussed in Theorem 10, $\Psi_0^{-1}(l, s)$ will be non-constant transfer function. Thus, $\Phi_{uu}^{-1}(p, r)(\omega)$ will be a non-constant transfer function. The case (q, s) can be shown similarly where r is a 1-hop neighbor of l in G^T..

(p, s): As p, s are 2-hop neighbors of l in G^T. Then, by (7.6) we have that $\Phi_{xx}^{-1}(p, l)$ and $\Phi_{xx}^{-1}(l, s)$ being real valued. Thus, $\Psi_0^{-1}(p, l)$ and $\Psi_0^{-1}(l, s)$ will be real valued transfer functions. As p, s are 4 hop neighbors, using (7.6) $\Phi_{xx}^{-1}(p, s) = 0$. This implies $\Psi_0^{-1}(p, s) = 0$. Then, $\Phi_{uu}^{-1}(\omega)(p, s) = -\Psi_0^{-1}(p, l)\Psi_0^{-1}(l, s)\Delta_l^{-1}$. We have $\Psi_0^{-1}(p, l)$, $\Psi_0^{-1}(l, s)$ and Δ_l^{-1} being real valued. Thus, $\Phi_{uu}^{-1}(p, s)(\omega)$ will be a real valued transfer functions. Therefore, $\angle\Phi_{uu}^{-1}(p, s)(\omega)$ will be a constant for all $\omega \in (-\pi, \pi]$. □

It follows from above proposition that only if the corresponding phase properties hold in $p - q - l - r - s$, then it is the only correct alignment as any other alignment will have non-2 hop neighbors as 4 hops away and hence will violate the constant phase argument. This verifies Lemma 5 and hence, Theorem 11 . □